THE UNIVERSITY OF MICHIGAN
JUN 24 1961
SCIENCE LIBRARY
PR
QH 301 B615 v.48 no.1-2

Vol. 48, Parts 1 and 2 June 1961

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITOR

J. DURBIN

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON

[*Issued June* 1961]

This volume of *Biometrika* is published with the co-operation of

F. N. DAVID — N. L. JOHNSON
M. J. R. HEALY — D. V. LINDLEY
R. L. PLACKETT

A volume containing about 500 pages will be published annually in two half-yearly issues, appearing in June and December.

Papers for publication should be sent to

PROFESSOR E. S. PEARSON
University College London, Gower Street, London, W.C. 1

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 50 copies of their papers free. Joint authors 25 copies each. Order forms for separates are sent to authors with proofs of their papers.

SUBSCRIPTIONS for 1961 (Volume 48)

The subscription price, payable in advance is:
£2. 14*s.* (or **$8.00**) net per volume *including packing and postage.*

Subscribers who buy the journal *Biometrika* for their own and not for general use may apply to the Secretary to be registered as **Personal Subscribers,** for whom the subscription price is **£2. 5*s.*** (payable only in sterling).

Cheques should be made payable to *Biometrika*, crossed 'a/c BIOMETRIKA TRUST' and sent to THE SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, W.C. 1, to whom all orders for Offprints, Index and Table Separates (see pages (iv) to (vi)) should also be sent. For particulars as to the availability of **Back Issues** please see inside back cover.

[Members of the Institute of Mathematical Statistics are reminded that they may subscribe to *Biometrika* through the Institute who will accept subscriptions to *Biometrika* at the special rate of $6.50 if paid to the Treasurer of the Institute before 1 March.]

Biometrika (1961), **48**, 1 *and* 2, *p.* 1
Printed in Great Britain

Studies in the history of probability and statistics

XI. Daniel Bernoulli on maximum likelihood

By M. G. KENDALL

Research Techniques Division, London School of Economics and Political Science

1. Almost as soon as the calculus of probabilities began to take a definite shape mathematicians were concerned with the use of probabilistic ideas in reconciling discrepant observations. James Bernoulli's *Ars Coniectandi* was published in 1713. Within 9 years we find Roger Cotes (1722), in a work on the estimation of errors in trigonometrical mensuration, discussing what would nowadays be described as an estimation problem in a plane. Let p, q, r, s be four different determinations of a point o, with weights P, Q, R, S which are inversely proportional to distance from o (*pondera reciproce proportionalia spatiis evagationum*). Put weights P at p, etc., and find their centre of gravity z. This, says Cotes, is the most probable site of o. (*Dico punctum z fore locum obiecti maxime probabilem, qui pro vero eius loco tutissime haberi potest.*) Cotes does not say why he thinks this is the most probable position or how he arrived at the rule.

2. According to Laplace this result of Cotes was not applied until Euler (1749) used it in some work on the irregularities in the motion of Saturn and Jupiter. Further attacks on the problem of a somewhat similar kind were employed by Mayer (1750) in a study of lunar libration and by Boscovich (1755) in measurements on the mean ellipticity of the earth. There was evidently a good deal of interest being taken in the combination of observations about the middle of the eighteenth century. The ideas, as was only natural, were often intuitive and sometimes obscurely expressed, but the fundamental questions seem to have been asked at quite an early stage. For example, Simpson (1757) refers to a current opinion that one good observation was as accurate as the arithmetic mean of a set, and although from that point onwards a series of writers argued for the arithmetic mean, Laplace (1774), in his first great memoir, was clearly aware that for some distributions of error there were better estimators such as the median.

3. Simpson (1756, 1757) was the first to introduce the concept of distribution of error and to consider continuous distributions. But like most of his contemporaries he regarded it as inevitable to impose two conditions: first, the distributions must be symmetrical; secondly, they must be finite in range. Lagrange reproduced Simpson's work without acknowledgement in a memoir published between 1770 and 1773, but Lagrange's contributions are more of analytical than of probabilistic interest.

4. Daniel Bernoulli was born in 1700 and lived to be 82. Throughout his productive life he made contributions to the theory of probability and although his mathematical methods are not now of much importance, the originality of his thinking on such matters as moral expectation entitles him to a permanent place among the founders of the subject. In particular, the memoir on maximum likelihood reproduced in the following pages is astonishingly in advance of its time. The author was 78 when it was published and it appears that he excogitated the basic ideas for himself without reference to previous writings. The memoir may, in actual fact, have been written rather earlier. Laplace's

article of 1774 refers to *manuscripts* of Bernoulli and Lagrange which he had heard of but not seen. An announcement of their existence, says Laplace sublimely, *reawakened* his interest in the subject. Laplace was 25 at the time.

5. I am much indebted to my colleague Mr C. G. Allen for the translations of the articles by Bernoulli and Euler which follow. They are, I felt, of sufficient interest to justify the publication of an English version, especially Bernoulli's. The reasoning is so clear that I can leave Daniel to tell his own story, but perhaps I may direct attention to two points:

(*a*) Influenced by the belief that an error distribution must have a finite range, Bernoulli runs into trouble with the parameter determining that range. He assumes a semi-circular distribution and lays down the peculiar condition that any distribution must be abrupt at its terminals. Once this is done, however, his formulation of maximum likelihood is clear and explicit and he derives what would nowadays be called the ML equations by differentiating the likelihood of the sample.

(*b*) In § 16 he is right on the verge of a principle of minimal variance. In comparing two methods of estimation he points out that one (the ML method) gives samples which are closer to the true value than the other.

6. The commentary by Euler seems to me of less value. He points out, correctly in my opinion, that the ML principle is arbitrary in the sense that there is no logical reason to believe that observations come from a generating system which gives them the greatest probability. (Bernoulli admits that his reasoning on this point is metaphysical, but at least he does reason about it.) Euler then goes on to propound principles which seem to me to be much more open to doubt than the one he is trying to replace. His examples at the end, in which he has to manoeuvre his error-range to avoid imaginary solutions, ends rather lamely with the conclusion that it doesn't matter much anyway. However, it is always of interest to read what a great mind has to offer on a subject. Nor should we forget, perhaps, that at the time of publication Euler himself was 71 and had been blind for 10 years.

REFERENCES

BOSCOVICH, R. G. (1755). (In Maire, C. and Boscovich.) *De litteraria expeditione per Pontificiam ditionem ad dimetiendos duos meridiani gradus.* Romae.

COTES, R. (1722). Aestimatio errorum in mixta mathesi, per variationes partium trianguli plani et spherici. *Opera Miscellanea*, Cantabrigiae.

EULER, L. (1749). *Pièce qui a remporté le prix de l'Académie Royale des Sciences en* 1748, *sur les inégalités du mouvement de Saturne et de Jupiter.* Paris.

LAGRANGE, J-L. (1770–3). Mémoire sur l'utilité de la méthode de prendre le milieu des résultats de plusieurs observations etc. *Miscellanea Taurinensia*, **5**, 167.

LAPLACE, P. S. (1774). Déterminer le milieu que l'on doit prendre entre trois observations données d'un même phénomène. *Mém. Acad. Paris* (*par divers savants*), **4**, 634.

MAYER, T. (1750). Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosmographische Nachrichten und Sammlung.*

SIMPSON, T. (1756). A letter...on the advantage of taking the mean of a number of observations in practical astronomy. *Phil. Trans.* **44**, 82.

SIMPSON, T. (1757). An attempt to show the advantage arising by taking the mean of a number of observations in practical astronomy. *Miscellaneous Tracts*, London.

The most probable choice between several discrepant observations and the formation therefrom of the most likely induction

By DANIEL BERNOULLI†

Translated by C. G. Allen

British Library of Political and Economic Science, London School of Economics and Political Science

1. Astronomers as a class are men of the most scrupulous sagacity; it is to them therefore that I choose to propound those doubts that I have sometimes entertained about the universally accepted rule for handling several slightly discrepant observations of the same event. By this rule the observations are added together and the sum divided by the number of observations; the quotient is then accepted as the true value of the required quantity, until better and more certain information is obtained. In this way, if the several observations can be considered as having, as it were, the same weight, the centre of gravity is accepted as the true position of the objects under investigation. This rule agrees with that used in the theory of probability when all errors of observation are considered equally likely.

2. But is it right to hold that the several observations are of the same weight or moment, or equally prone to any and every error? Are errors of some degrees as easy to make as others of as many minutes? Is there everywhere the same probability? Such an assertion would be quite absurd, which is undoubtedly the reason why astronomers prefer to reject completely observations which they judge to be too wide of the truth, while retaining the rest and, indeed, assigning to them the same reliability. This practice makes it more than clear that they are far from assigning the same validity to each of the observations they have made, for they reject some in their entirety, while in the case of others they not only retain them all but, moreover, treat them alike. I see no way of drawing a dividing line between those that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others. Nevertheless, I do not condemn in every case the principle of rejecting one or other of the observations, indeed I approve it, whenever in the course of observation an accident occurs which in itself raises an immediate scruple in the mind of the observer, before he has considered the event and compared it with the other observations. If there is no such reason for dissatisfaction I think each and every observation should be admitted whatever its quality, as long as the observer is conscious that he has taken every care.

3. Let us compare the observer with an archer aiming his arrows at a set mark with all the care that he can muster. Let his mark be a continuous vertical line so that only deviations in a horizontal direction are taken into account; let the line be supposed to be drawn in the middle of a vertical plane erected perpendicular to the axis of vision, and let the whole of the plane on either side be divided into narrow vertical bands of equal width. Now if the arrow be loosed several times, and for each shot the point of impact be examined and its distance from the vertical mark noted on a sheet, though the outcome cannot in the least be exactly predicted, yet there are many assumptions that can reasonably be made and

† This memoir and the following commentary by Euler appeared in Latin in the memoirs of the Academy of St Petersburg, *Acta Acad. Petrop.* (1777), pp. 3–33. A photostat copy has been deposited in the library of the Royal Statistical Society.

which can be useful to our inquiry, provided all the errors are such as may easily be in one direction as the other, and their outcome is quite uncertain, being decided only as it were by unavoidable chance. In astronomy, likewise, anything which admits of correction *a priori* is not reckoned as an error. When all those corrections have been made which theory enjoins, any further correction which is necessary in order to reconcile the several slightly discrepant observations which differ slightly from each other is a matter solely for the theory of probability. What in particular happens in the course of observation, *ex hypothesi* we scarcely know, but this very ignorance will be the refuge to which we are forced to flee when we take our stand on what is not truest but most likely, not certain but most probable (*non verissimum sed verisimillimum, non certum sed probabilissimum*), as the theory of probability teaches. Whether that is always and everywhere identical with the usually accepted arithmetical mean may reasonably be doubted.

4. Errors, which are unavoidable in observation, may indeed affect individual observations; nevertheless, any given observation has its own rights and could not be impugned if it were the only one that had been made. Any observation must therefore be in itself sound and good, and no-one ought to assign any other value than that ascertained thereby; but since they are mutually contradictory, a value has to be assigned to the whole complex of observations without touching the parts. In this way a definite error is attributed to the individual observations; but I think that of all the innumerable ways of dealing with errors of observation one should choose the one that has the highest degree of probability for the complex of observations as a whole.

The rule which I here propound will be accepted by all, provided that the degree of probability in respect of a given observation can be defined in terms of a point which is assumed to be true. I freely admit that this last condition has not been definitely met; at the same time I am convinced that all things are not equally uncertain and that better results can be got than can be expected from the commonly accepted rule. Let us see if certain assumptions should not properly be made in this argument which contribute something to a higher probability. I will begin the examination with some general considerations.

5. If the archer whom I mentioned in § 3 makes innumerable shots, all with the utmost possible care, the arrows will strike sometimes the first band next to the mark, sometimes the second, sometimes the third and so on, and this is to be understood equally of either side whether left or right. Now is it not self-evident that the hits must be assumed to be thicker and more numerous on any given band the nearer this is to the mark? If all the places on the vertical plane, whatever their distance from the mark, were equally liable to be hit, the most skilful shot would have no advantage over a blind man. That, however, is the tacit assertion of those who use the common rule in estimating the value of various discrepant observations, when they treat them all indiscriminately. In this way, therefore, the degree of probability of any given deviation could be determined to some extent *a posteriori*, since there is no doubt that, for a large number of shots, the probability is proportional to the number of shots which hit a band situated at a given distance from the mark.

Moreover, there is no doubt that the greatest deviation has its limits which are never exceeded and which indeed are narrowed by the experience and skill of the observer. Beyond these limits all probability is zero; from the limits towards the mark in the centre the probability increases and will be greatest at the mark itself.

6. The foregoing give some idea of a scale of probabilities for all deviations, such as each observer should form for himself. It will not be absolutely exact, but it will suit the nature

of the inquiry well enough. The mark set up is, as it were, the centre of forces to which the observers are drawn; but these efforts are opposed by innumerable imperfections and other tiny hidden obstacles which may produce in the observations small chance errors. Some of these will be in the same direction and will be cumulative, others will cancel out, according as the observer is more or less lucky. From this it may be understood that there is some relation between the errors which occur and the actual true position of the centre of forces; for another position of the mark the outcome of chance would be estimated differently. So we arrive at the particular problem of determining the most probable position of the mark from a knowledge of the positions of some of the hits. It follows from what we have adduced that one should think above all of a scale (*scala*) between the various distances from the centre of forces and the corresponding probabilities. Vague as is the determination of this scale, it seems to be subject to various axioms which we have only to satisfy to be in a better case than if we suppose every deviation, whatever its magnitude, to occur with equal ease and therefore to have equal probability. Let us suppose a straight line in which there are disposed various points, which indicate of course the results of different observations. Let there be marked on this line some intermediate point which is taken as the true position to be determined. Let perpendiculars expressing the probability appropriate to a given point be erected. If now a curve is drawn through the ends of the several perpendiculars this will be the scale of the probabilities of which we are speaking.

7. If this is accepted, I think the following assumptions about the scale of probabilities can hardly be denied.

(*a*) Inasmuch as deviations from the true intermediate point are equally easy in both directions, the scale will have two perfectly similar and equal branches.

(*b*) Observations will certainly be more numerous and indeed more probable near to the centre of forces; at the same time they will be less numerous in proportion to their distance from that centre. The scale therefore on both sides approaches the straight line on which we supposed the observed points to be placed.

(*c*) The degree of probability will be greatest in the middle where we suppose the centre of forces to be located, and the tangent to the scale for this point will be parallel to the aforesaid straight line.

(*d*) If it is true, as I suppose, that even the least-favoured observations have their limits, best fixed by the observer himself, it follows that the scale, if correctly arranged, will meet the line of the observations at the limits themselves. For at both extremes all probability vanishes and a greater error is impossible.

(*e*) Finally, the maximum deviations on either side are reckoned to be a sort of boundary between what can happen and what cannot. The last part, therefore, of the scale, on either side, should approach steeply the line on which the observations are sited, and the tangents at the extreme points will be almost perpendicular to that line. The scale itself will thus indicate that it is scarcely possible to pass beyond the supposed limits. Not that this condition should be applied in all its rigour if, that is, one does not fix the limits of error over-dogmatically.

8. If we now construct a semi-ellipse of any parameter on the line representing the whole field of possible deviations as its axis, this will certainly satisfy the foregoing conditions quite well. The parameter of the ellipse is arbitrary, since we are concerned only with the proportion between the probabilities of any given deviation. However elongated or compressed the ellipse may be, provided it is constructed on the same axis, it will perform the

same function; which shows that we have no reason to be anxious about an accurate description of the scale. In fact we can even use a circle, not because it is proved to be the true scale by mathematical reasoning, but because it is nearer the truth than an infinite straight line parallel to the axis, which supposes that the several observations are of equal weight and probability, however distant from the true position. This circular scale also lends itself best to numerical calculations; meanwhile it is worth observing in advance that both hypotheses come to the same whenever the several observations are considered to be infinitely small. They also agree if the radius of the auxiliary circle is supposed to be infinitely large, as if no limits were set to the deviations. Thus if the deviation of an observation from the true position is thought of as the sine of a circular arc, the probability of that observation will be the cosine of the same arc. Let the auxiliary semicircle, which I have just described, be called *the controlling semicircle* (*moderator*). Where the centre of this semicircle is located, the true position, which fits the observations best, is to be fixed. Admittedly our hypothesis is, to some extent, precarious, but it is certainly to be preferred to the common one, and will not be hazardous to those who understand it, since the result that they will arrive at will always have a higher probability than if they had adhered to the common method. When by the nature of the case a certain decision cannot be reached, there is no other course than to prefer the more probable to the less probable.

9. I will illustrate this line of argument by a trivial example. The particular problem is the reconciliation of discrepant observations; it is therefore a question of difference of observations. Now if a dice-thrower makes three throws with one die so that the second exceeds the first by one and the third exceeds the second by two, the throws may arise in three ways, viz. 1, 2, 4 or 2, 3, 5 or 3, 4, 6. None of these throws is to be preferred to the other two, for each is in itself equally probable. If you prefer the one in the middle, viz. 2, 3, 5, the preference is illogical. The same sort of thing happens if you choose to consider observations which, so far as you are concerned, are accidental, whether they are astronomical or of some other kind, as equally probable. Now suppose the thrower produces the same result by throwing a pair of dice three times. There will then be eight different ways in which he would obtain this result, viz. 2, 3, 5; 3, 4, 6; 4, 5, 7; 5, 6, 8; 6, 7, 9; 7, 8, 10; 8, 9, 11 and 9, 10, 12. But they are far from being all equally probable. It is well known that the respective probabilities are proportional to the numbers 8, 30, 72, 100, 120, 80, 40 and 12. From this known scale I have better right to conclude that the fifth set has happened than that any other has, because it has the highest probability; and so the three throws of a pair of dice will have been 6, 7 and 9. No-one, however, will deny that the first set 2, 3 and 5 might possibly have happened, even though it has only a fifteenth part of the probability corresponding to the fifth set. Forced to choose, I simply choose what is most probable. Although this example does not quite square with our argument, it makes clear what contribution the investigation of probabilities can make to the determination of cases. Now I will come more to grips with the actual problem.

10. First of all, I would have every observer ponder thoroughly in his own mind and judge what is the greatest error which he is morally certain (though he should call down the wrath of heaven) he will never exceed however often he repeats the observation. He must be his own judge of his dexterity and not err on the side of severity or indulgence. Not that it matters very much whether the judgement he passes in this matter is fitting or somewhat flighty. Then let him make the radius of the *controlling circle* equal to the aforementioned greatest error; let this radius be r and hence the width of the whole doubtful field $= 2r$.

If you desire a rule on this matter common to all observers, I recommend you to suit your judgement to the actual observations that you have made: if you double the distance between the two extreme observations, you can use it, I think, safely enough as the diameter of the controlling circle, or, what comes to the same thing, if you make the radius equal to the difference between the two extreme observations. Indeed, it will be sufficient to increase this difference by half to form the diameter of the circle if several observations have been made; my own practice is to double it for three or four observations, and to increase it by half for more. Lest this uncertainty offend any one, it is as well to note that if we were to make our controlling semicircle infinite we should then coincide with the generally accepted rule of the arithmetical mean; but if we were to diminish the circle as much as possible without contradiction, we should obtain the mean between the two extreme observations, which as a rule for several observations I have found to be less often wrong than I thought before I investigated the matter.

11. After all these preliminaries it remains to determine the position of the controlling circle, since it is at the centre of this circle that the several observations should be deemed to be, as it were, concentrated. The aforesaid position is deduced from the fact that the whole complex of observations would occur more easily, and therefore more probably, for this location than for any other position of the circle. We shall have the true degree of probability for the whole complex of observations if we note the probability corresponding to the several observations that have been carried out and multiply all the probabilities by each other, just as we did in § 9. Then the product of the multiplication is to be differentiated and the differential put = 0. In this way we shall obtain an equation whose root will give the distance of the centre from any given point.

Put the radius of the controlling circle = r; the smallest observation = A; the second $A+a$; the third $A+b$; the fourth $A+c$, and so on; the distance of the centre of the controlling semicircle from the smallest observation = x, so that $A+x$ will denote the quantity which is most probably to be assumed on the basis of all the observations. By our hypothesis the probability for the first observation alone is to be expressed by $\sqrt{\{r^2-x^2\}}$; for the second observation by $\sqrt{\{r^2-(x-a)^2\}}$; for the third by $\sqrt{\{r^2-(x-b)^2\}}$; for the fourth by $\sqrt{\{r^2-(x-c)^2\}}$ and so on. Then I would have the several probabilities multiplied together according to the rules of the theory of probability, which gives

$$\sqrt{\{r^2-x^2\}} \times \sqrt{\{r^2-(x-a)^2\}} \times \sqrt{\{r^2-(x-b)^2\}} \times \sqrt{\{r^2-(x-c)^2\}} \times \ldots.$$

Finally, if the differential of this product is put = 0, the equation, by virtue of our hypotheses, gives the required value x as having the highest probability. As, however, the aforesaid quantity is to be brought to its maximum value, it is obvious that its square will simultaneously be brought to the same state. So we can use, for ease of calculation, a formula which is composed entirely of rational terms, viz.

$$(r^2-x^2) \times \{r^2-(x-a)^2\} \times \{r^2-(x-b)^2\} \times \{r^2-(x-c)^2\} \times \ldots$$

and the differential is once more put = 0. For the rest, as many factors are to be taken as there were observations.

12. If a single observation was made, we must accept the observation as true. Now this is shown by our hypothesis. If only the first factor r^2-x^2 is taken, we shall have $-2xdx = 0$ or $x = 0$ and consequently $A+x = A$. So in this case our hypothesis agrees with the common one.

If two observations have been made, A and $A+a$, two factors are to be taken, namely

$$\{r^2-x^2\}\times\{r^2-(x-a)^2\} \quad \text{or} \quad r^4-2r^2x^2+x^4+2ar^2x-a^2r^2+2ax^3\times a^2x^2,$$

the differential of which

$$=-4r^2x\,dx+4x^3dx+2ar^2dx-6ax^2dx+2a^2x\,dx=0 \quad \text{or} \quad 2x^3-3ax^2-2r^2x+a^2x+ar^2=0.$$

The only useful root which this equation gives is $x=\frac{1}{2}a$, and $A+x=A+\frac{1}{2}a$. This also is the teaching of the common hypothesis. This agreement holds whatever be the radius of the controlling circle, a fact which shows clearly enough, in the case of several observations, that the size of our controlling circle in an enterprise of this sort need not be strictly exact, and one should not expect it to be. What is awkward—and I do not conceal it—is that for several observations a very long calculation is required, and so I hardly dare propose more than general discussions of these cases. Let me at least expound the theory of three observations, which is of the highest importance.

13. When we have three observations to deal with, viz. A; $A+a$ and $A+b$, we shall have three factors

$$\{r^2-x^2\}\times\{r^2-(x-a)^2\}\times\{r^2-(x-b)^2\},$$

for which we have to find the maximum value. If now these factors are actually multiplied together we shall obtain

$$\begin{gathered}
r^6+2ar^4x-3r^4x^2-4ar^2x^3+3r^2x^4+2ax^5-x^6\\
-a^2r^4-2ab^2r^2x+2b^2r^2x^2+2ab^2x^3-b^2x^4+2bx^5\\
-b^2r^4+2br^4x-a^2b^2x^2-4br^2x^3-4abx^4\\
+a^2b^2r^2-2a^2br^2x+4abr^2x^2+2a^2bx^3-a^2x^4\\
+2a^2r^2x^2.
\end{gathered}$$

If this expression is differentiated, and then after division by dx is put $=0$ to obtain the maximum value, the following general equation for any three observations whatsoever will result

$$\begin{gathered}
2ar^4-6r^4x-12ar^2x^2+12r^2x^3+10ax^4-6x^5\\
-2ab^2r^2+4b^2r^2x+6ab^2x^2-4b^2x^3+10bx^4\\
+2br^4-2a^2b^2x-12br^2x^2-16abx^3\\
-2a^2br^2+8abr^2x+6a^2bx^2-4a^2x^3\\
+4a^2r^2x=0.
\end{gathered}$$

The root of this equation, which is indeed of the fifth degree and consists of twenty terms, gives the distance of the centre of the controlling circle from the first observation, and the quantity $A+x$ gives the value which is most probably to be deduced from the three observations which have been made.

14. Unless the force of our fundamental arguments has been most attentively weighed there will be few perhaps who will see any relation whatever between the enormous equation and what seems to be a very simple question; for the common answer is $x=\frac{1}{3}(a+b)$. Nevertheless, our equation corresponds well enough to notions which crop up elsewhere, some of which I will now expound.

(*a*) If the radius of the controlling circle is supposed to be infinite compared with a and b, all terms are to be rejected except those in which r rises to the highest power, in which case

our equation is reduced to this very simple one $2ar^4+2br^4-6r^4x=0$ or $x=\frac{1}{3}(a+b)$. So the common rule is contained in our equation. If, however, our definition set out in § 10 is considered, it will be obvious how unfitting is the hypothesis of an infinite radius and how manifestly some more suitable one could be substituted for it.

(*b*) If we put $b=2a$, it is obvious that $x=a$ whatever value is given to the radius r, and that too will be common to both theories. Let us see therefore what our equation shows for this case. Substituting for b the equation becomes

$$6ar^4-6r^4x-36ar^2x^2+12r^2x^3+30ax^4-6x^5$$
$$-12a^3r^2+36a^2r^2x+36a^3x^2-52a^2x^3$$
$$-8a^4x=0.$$

Now this equation, whatever be the value of r, is satisfied by $x=a$, which the nature of the case demands.

(*c*) If $b=-a$, x must equal 0 whatever be the value of r. This too is beautifully shown by our equation, which now becomes

$$-6r^4x+12r^2x^3-6x^5$$
$$-2a^4x+8a^2x^3=0.$$

A glance will show that the useful root is $x=0$.

15. This and other similar corollaries sufficiently confirm the real connexion of our fundamental arguments with the question under discussion, however enormous the equation we have found may seem in so simple an inquiry. I proceed to examples in which the radius of the controlling circle is neither infinite nor indifferent, which is where practically all cases belong. In these examples our new theory always produces a different result from the common one; and the more the intermediate observation approaches either extreme, the greater the difference. It is on the discussion of these cases that the matter hinges, so we must have recourse to purely numerical examples.

Example 1. Let us assume three observations

$$A;\quad A+0{\cdot}2\quad\text{and}\quad A+1,$$

so that

$$a=0{\cdot}2\quad\text{and}\quad b=1$$

and let the value to be assumed as most likely from these three observations be $A+x$. The common rule gives $x=0{\cdot}4$. Let us see the new one which to my mind is more probable, and let us put $r=1$ (cf. § 10). The following purely numerical equation results

$$1{\cdot}92-0{\cdot}32x-12{\cdot}96x^2+4{\cdot}64x^3+12x^4-6x^5=0.$$

the solution of which is approximately $x=0{\cdot}4427$, which exceeds the commonly accepted value by more than a tenth. This marked excess is due to the fact that the middle observation is much nearer to the first than to the third. From this it is easily deduced that the excess will be changed to a defect if the middle observation is nearer to the third than to the first, and that the nearer the middle observation is to the mean between the two extreme observations, the smaller will be this defect. To test this conjecture I retain the other values and change only the middle observation, as follows.

Example 2. Let a now $=0{\cdot}56$, and as before $r=b=1$. By the commonly accepted rule we shall have $x=0{\cdot}52$. Let us see what happens with ours. The equation of § 13 gives the following numerical equation

$$1{\cdot}3728+3{\cdot}1072x-13{\cdot}4784x^2-2{\cdot}2144x^3+15{\cdot}6x^4-6x^5=0$$

which is approximately satisfied by $x = 0{\cdot}5128$. In accordance with our principles, the value of x is less than the arithmetical mean which is usually accepted, but the difference between the two is now quite small, viz. 0·0072, exactly as I had anticipated would be the case. Hence it can also be seen that the greatest difference between the two estimates occurs when it so happens that two observations exactly coincide and only the third diverges. There are two cases, viz. when $a = 0$ and when $a = b$. I will expound the result in each case.

Example 3. Put $a = 0$, leaving the remaining denominations unaltered. Dividing by $2b - 2x$ we have the following numerical equation

$$1 - 6x^2 - 2x^3 + 3x^4 = 0,$$

which is approximately satisfied by $x = 0{\cdot}3977$, whereas the value of x obtained from the common rule is $x = 0{\cdot}3333$. The former exceeds the latter by 0·0644. If, however, we put $a = b$ and divide by $2x$, the following equation results

$$4 - 6x - 6x^2 - 10x^3 - 3x^4 = 0.$$

This is approximately satisfied by $x = 0{\cdot}6022$, while the common value is 0·6666. So the difference between the two is once more 0·0644, but this time our new value is less than the common one, whereas previously it was greater. It is clear from this that our method takes better aim at a certain intermediate point than does the common method. Evidence of this sort does much to commend the method that I propose, and I will go a little more closely into this consideration, if so be that an *argumentum ad hominem* may be accepted in a matter which does not admit of mathematical demonstration.

16. If we combine the two cases in example 3, and suppose that six observations have been made, viz. $A, A, A+b$ and $A+b, A+b, A$, it is obvious that three observations support the value A and the same number the value $A+b$. We see by § 12 that in this case both methods give the required mean value as $A + \frac{1}{2}b$, or for example 3, $A + 0{\cdot}5$; or, omitting the constant quantity A, simply 0·5. This value, derived from the six observations combined, will not be doubted by anyone. Now let us divide these six observations into two other triads, namely $A, A, A+1$ and $A+1, A+1, A$. In this case, rejecting once more the quantity A, the commonly accepted rule gives for the first triad $0{\cdot}\dot{3}$ and for the second $0{\cdot}\dot{6}$, both differing, the first by defect and the second by excess, by $0{\cdot}1\dot{6}$ from the mean 0·5. So for either triad of observations taken separately the common theory involves an error of $0{\cdot}1\dot{6}$, while ours involves an error of 0·1022, which is notably smaller. A great deal more evidence of this kind could be adduced to give further support to our fundamental argument; but I am afraid I should appear immoderate if I went on extending something which cannot be settled with certainty and absolute perfection. We have no higher aim than to be able to distinguish what is more probable from what is less.

17. Such further perfection as we may reasonably expect will consist in a stricter and more accurate determination of the controlling scale and its width. I will add a few further comments on this topic. It is obvious from the foregoing considerations that our estimates are not so very different from the commonly accepted rule: so it is a question of a certain correction which this rule appears to allow. This correction is provided by the actual divergences of the observations from the required true point, since they can be so arranged, for any given width of the controlling scale, as to make the most probable fit with this point. But for my part I can see no way of strictly determining the width of the aforesaid scale except that which I mentioned in § 10. If an observer, through undue mistrust of his own powers, enlarges the

dimensions of the controlling semicircle excessively, it will not give all the help it might, but what it gives will be more certain; if on the other hand he contracts the scale unduly, other things being equal he will arrive at a correction which is a little greater and somewhat less probable. Prudence seems to be as necessary here as sharp-sightedness. Should you wish to use the observations that have actually been made as a basis for an *a posteriori* estimate of the width of the controlling scale to be applied, it will be prudent to weigh in your own mind whether one should consider the observations to have turned out luckily or not. The more you assign to good luck, the less you can attribute to the skill applied in observing, and the larger accordingly will be the controlling circle which you will apply. In § 13 I assumed $r = b$; in other words the radius of the controlling circle equalled the distance between the two extreme observations. I admit, however, on better reflexion that this size of radius seems to me to argue somewhat excessive confidence; it would be safer certainly in future to put $r = \frac{3}{2}b$ or even $r = 2b$. If so, the correction would come out notably smaller but all the more certain and trustworthy.

18. If there is any validity in our principles, though they are metaphysical rather than mathematical, we may justly conclude therefrom that one should seldom if ever reject an observation, and never without the utmost circumspection. I have already given my opinion on this subject in § 2. The whole complex of observations is simply a chance event modified and confined within certain limits by the skill of the observer. It may well happen, though very rarely, that of three observations two are miraculously identical, while the third by ill luck is very wide of the other two. But if this happens to me and I am certain that I have not unduly contracted the limits of maximum possible error or shown undue confidence in my skill, I should not hesitate to refer the examination of the whole case to our principles and form my estimate from them. Only the observer must give the same attention to each of the observations. I should like them all treated equally.

19. The only remaining caution refers to the controlling scale which I have applied. We have taken a semicircle as answering sufficiently the conditions set out in § 7 and at the same time most suited to the calculations that have to be carried out. Meanwhile it is worthy of note that there are other infinite curves which undoubtedly lead to the same equation as I set out in § 13. In § 11 we made the probabilities, for a circular scale, proportional respectively to the perpendiculars

$$\sqrt{(r^2-x^2)}; \quad \sqrt{\{r^2-(x-a)^2\}}; \quad \sqrt{\{r^2-(x-b)^2\}}.$$

Now if instead of a semicircle we suppose a parabola (*arcum parabolicum*) constructed on the line $2r$, with its axis passing perpendicularly through the middle of this line, then keeping the same notation, we shall have perpendiculars, or the corresponding probabilities expressed by them,

$$\frac{\rho}{r^2}(r^2-x^2), \quad \frac{\rho}{r^2}\{r^2-(a-x)^2\}, \quad \frac{\rho}{r^2}\{r^2-(b-x)^2\}, \quad \text{etc.},$$

where the new letter ρ denotes the longest perpendicular at the abscissa $x = 0$. Now since the factor ρ/r^2 is common to all the terms we can simply substitute unity for this factor when we have brought the product of all the several probabilities to a maximum. It follows from this that the parameter of the parabola is always arbitrary. I also pointed out in the aforementioned § 11 that if this product has been brought to its maximum all its powers will at the same time be maximized or minimized. It is obvious from this that both scales, the parabolic and the circular, lead to the same required value of x. Furthermore, it is

evident that innumerable other scales fulfil the same function; they will all have this property, that from their peak they approach in either direction the line $2r$, on which the several observations are necessarily supposed to lie, and intersect it. Therefore all scales of this sort achieve our aim, and we need not be too pedantic in this matter, since we are content to strive for something better if not for the best.

20. Finally, as regards the awkward, not to say monstrous, form of our fundamental equation set out in § 13, we can mend the awkwardness somewhat; for I express the useful root as approximately

$$x = \frac{a+b}{3} + \frac{2a^3 - 3a^2b - 3ab^2 + 2b^3}{27r^2}.$$

The first term is none other than the common arithmetical mean for three observations, the second indicates approximately the further correction required by our principles. This root indeed will agree all the more accurately with the equation of § 13, the greater is assumed to be the width of the controlling scale indicated by $2r$. Far be it from us, however, to increase the value of the letter r unnecessarily merely to make calculation easier, for every useless increase takes away a little from the amount of our correction. Nor would it be less dangerous to attribute too much to one's powers of observation and so shorten the radius r unjustifiably. 'There are fixed bounds, outside of which justice cannot exist': cf. § 10. Our principles themselves show that it is impossible for r to be less than $\frac{1}{2}b$, since this involves the manifest contradiction of positing as impossible something which is supposed to have actually happened. I have not concealed, however, the somewhat free assumptions that have been made in the course of our argument; but I should not have thought that all our methods of judging the observations that have been made ought to be rejected on that account. Of this at least I am convinced, that *the common rule for three observations gives somewhat too small a result when* $a < \frac{1}{2}b$ *and too large a result if* $a > \frac{1}{2}b$, *and cannot ever be applied with greater certainty than when the intermediate observation is approximately equidistant from the two extremes.* Secondly, I think it probable that our equation in § 13 gives a safer and better determination of the position to be selected, provided the radius of the controlling circle is not rashly diminished beyond the limits which the powers of the observer permit: cf. § 17. The question that I have dealt with is properly this: given three or more shots of an arrow marked on a straight line, to determine the most probable position of the point at which the archer was aiming. But any and every observer who understands these things will form for himself criteria which will answer his purpose, according to the nature of the material (*argumento*) which he has to hand, provided he makes cautious use of the rules derived from the theory of combinations.

Recapitulation. By its very nature our problem is indeterminate, inasmuch as it depends on the practice, experience, and skill of the observer, on the precision of the instruments, on the keenness of the senses, in short on countless circumstances which may be more or less favourable. Account will be taken of all these things in assuming the width of the field of possible deviations; on this subject I have given my opinion, with all circumspection. Secondly, one has to examine the casual working of chance in favour of any given deviation (lit. the working of the casual chance which favours any deviation), since it is advantageous if any given deviation is assigned the probability which from the nature of the case fits it.† To be sure, this scale of probabilities remains in its turn uncertain and undetermined, should

† Reading *cuiuis aberrationi* for the *cuiuis aberratione* of the text.

an accurate one be desired, but displays, nevertheless, by the very nature of the case, several properties; and if these are satisfied, it may be considered to be sufficiently known, as I learned from several experiments. So a method comes to light of expressing in accordance with the proven precepts of the theory of probability the absolute probability appropriate to any given system of observations for any assumed location of that system. It only remains then to select that location of the system in question which enjoys the highest probability. It certainly seems extraordinary to me that the algebraic equation defining this location, which is so far-fetched and rises to the fifth degree for only three observations, which is expressed in a very large number of terms and is deduced from principles never used before, nevertheless, from whatever point it is examined, gives rise to nothing which is in the least displeasing, still less leads to any absurd result. The upshot of the calculations in any example is little different from that which is indicated by the common method, provided one does not recklessly jump at the precepts which I have laid down. Where the comparison of three given observations shows that the middle one is approximately equidistant from the extremes, we shall adhere without hesitation to the common rule; but if the two intervals are notably unequal, I think it is better to have recourse to our theory, provided one follows the precepts I have set out and exercises the greatest prudence in fixing just bounds to the field of possible deviations. All this I should wish to have weighed in the balance of metaphysics rather than mathematics. Those who are most shocked by our principles will have nothing further to contradict if only they make the field of possible deviations as large as possible.

Observations on the foregoing dissertation of Bernoulli

By L. EULER

1. The question which our distinguished friend Bernoulli handles here is one of no little moment, namely, how an unknown quantity should be derived from several observations which vary slightly from each other. To make the nature of the question easier to discern clearly, let us suppose that the elevation of the pole star at some place or other has to be discovered and that the observations made to this end have the following different values:

$$\Pi + a, \quad \Pi + b, \quad \Pi + c, \quad \Pi + d, \quad \text{etc.},$$

where the letters a, b, c, d, etc., are taken to be expressed in seconds. From these the true elevation of the pole at this place, $\Pi + x$, is to be deduced. Generally this quantity x is obtained by taking the arithmetic mean of all the quantities a, b, c, d, etc. Hence if the number of observations $= n$, $x = (a+b+c+d+\text{etc.})/n$.

2. In this rule it is obviously assumed that all observations are of the same degree of goodness. For if some were more exact than others, account ought to be taken of this distinction in the computation. Now although there is no apparent reason in the circumstances why one of these observations should be accorded a greater value than the rest, nevertheless, the learned author observes that these observations ought to be awarded a higher degree of goodness the nearer they approach to the truth, just as that class of observations which is thought to depart too far from the truth is usually completely rejected. The whole business therefore amounts to this: to show how the degree of goodness appropriate to the several observations is to be estimated.

3. According to the view of the distinguished author, it will be convenient to consider the deviation of each observation from the truth as already known. This will be $x-a$ for the first observation, $x-b$ for the second, $x-c$ for the third, etc., but the defect of each observation should be estimated not so much from these differences as from their squares, since the defect itself is to be reckoned as the same whether the observation errs by excess or defect. Hence if some observation agrees perfectly with the truth, its defect will be zero. If therefore the degree of goodness of this observation is indicated by r^2, it is obvious that the degree of goodness of the first observation must be indicated by $r^2-(x-a)^2$, that of the second by $r^2-(x-b)^2$, of the third by $r^2-(x-c)^2$ and so on, the value of r being such that for an observation which is to be all but rejected the degree of goodness vanishes. If we assume that this happens in the observation which gives $\Pi+u$, then since the degree of goodness of this would be $r^2-(x-u)^2$, it must be laid down in all cases that $r^2=(x-u)^2$.

4. Having established these conclusions concerning the degree of goodness of each observation, the distinguished author appeals to the following principle, for which indeed he gives no reason: that the product of all the formulae expressing the degrees of goodness of the several observations should be allotted a maximum value. On this principle therefore he bids one differentiate this product and equate the differential with nought, since this equation will then give the true value of x. This he illustrates with some examples based on sets of three observations, deriving therefrom values of x which seem to be quite in conformity with the truth.

5. This principle for only three observations led to an equation of the fifth degree, whose root x had to be found; and anyone who wished to apply the principle to four observations would arrive at an equation of the seventh degree. Five observations would lead to one of the ninth degree and so on. It is thus abundantly evident that this method cannot possibly be used where there are several observations, and this is in fact candidly conceded by the distinguished author, who presents the whole dissertation as a purely metaphysical speculation.

6. As, however, the distinguished author has not supported this principle of the maximum by any proof, he will not take it amiss if I propound certain doubts about it. If we assume that among the observations in question there is one that should be almost rejected, whose degree of goodness would accordingly be as small as possible, it is evident that the product of all the formulae mentioned would in fact be reduced to nothing, so that it could not possibly be considered as a maximum, no matter how great it might be, were that observation omitted. Now the principles of the theory of probability make it abundantly clear that the value of the unknown quantity x should come out the same whether an observation such as this, which has no goodness at all, is introduced into the calculation or totally rejected.

7. I do not think that it is necessary in this question to have recourse to the principle of the maximum, since the undoubted precepts of the theory of probability are quite sufficient to resolve all questions of this kind. If the first observation, which gave $\Pi+a$, is assigned the amount or degree of goodness (*pretium seu gradum bonitatis*) α, the second β, the third γ, the unknown quantity x is given by the rules of this theory thus:

$$x=\frac{a\alpha+b\beta+c\gamma+d\delta+\text{etc.}}{\alpha+\beta+\gamma+\delta+\text{etc.}}.$$

Hence
$$\alpha(x-a)+\beta(x-b)+\gamma(x-c)+\delta(x-d)+\text{etc.}=0.$$

Now it is clear that if all the grades of goodness were equal and the number of observations were n, we should have $x = (a+b+c+d+\text{etc.})/n$, as required by the common rule. From which it follows that different values may emerge for the unknown quantity x to the extent that the degrees of goodness differ.

8. Since therefore, as the distinguished author himself states, the grades of goodness indicated by the letters $\alpha, \beta, \gamma, \delta$ are

$$\alpha = r^2-(x-a)^2, \quad \beta = r^2-(x-b)^2,$$
$$\gamma = r^2-(x-c)^2, \quad \delta = r^2-(x-d)^2, \quad \text{etc.},$$

the equation we have found becomes

$$r^2(x-a)+r^2(x-b)+r^2(x-c)$$
$$-(x-a)^3-(x-b)^3-(x-c)^3\,\text{etc.} = 0.$$

Hence if the number of observations $= n$ and we put for brevity's sake

$$a+b+c+d+\text{etc.} = A,$$
$$a^2+b^2+c^2+d^2+\text{etc.} = B,$$
$$a^3+b^3+c^3+d^3+\text{etc.} = C,$$

that equation is reduced to the following fairly simple form

$$nr^2x-Ar^2-nx^3+3Ax^2-3Bx+C = 0.$$

Thus we arrive at a cubic equation, from which the unknown x can easily be found, whatever the number of observations n.

9. If we regard the quantity r as infinite, which is the case when all the observations are assigned the same degree of goodness, then we may neglect all the other terms and directly deduce from this equation the following

$$x = \frac{A}{n} = \frac{a+b+c+d+\text{etc.}}{n}$$

just as is required by the rule which is commonly adopted. If we designate this value by the letter p, and substitute Π for $\Pi+p$ in the observations themselves, we shall have to diminish the several numbers a, b, c, d, etc., by the same quantity p, and thus the sum of them all, for which we put A, will equal 0. To avoid, however, the introduction of new letters into the calculation at this point we can from the beginning so constitute the quantity Π that if the values of the several observations are given as $\Pi+a$, $\Pi+b$, $\Pi+c$, $\Pi+d$, etc., the sum of the letters $a+b+c+d+\text{etc.} = 0$. Then to discover the quantity x we shall have the following much simpler equation

$$nx^3-nr^2x+3Bx-C = 0,$$

from which would follow, if r were infinite, $x = 0$. It is clear from this that if this equation has several real roots, the smallest should be taken as x, so that the required true value will be $\Pi+x$.

10. This same question can, however, be referred even to a quadratic equation by introducing the sort of observation which after weighing all the circumstances we decide should be totally rejected. Let such an observation be $\Pi+u$, and since *ex hypothesi* its degree of

goodness $r^2-(x-u)^2=0$, $r^2=(x-u)^2$. The introduction of this value in the last equation that we found produces the following form

$$2nux^2-nu^2x+3Bx-C=0.$$

It will be convenient to regard the term $-nu^2x$ in this equation as the greatest, so that the equation can be expressed as follows

$$x(nu^2-3B-2nux)=-C$$

from which follows

$$x=\frac{-C}{nu^2-3B-2nux},$$

where by substituting for x the value just obtained we get the following continued fraction

$$x=\frac{-C}{nu^2-3B+}\ \frac{2nuC}{nu^2-3B+}\ \frac{2nuC}{nu^2-3B+},\quad \text{etc.},$$

a form which will soon give the true value of x itself.

11. Since the distinguished author has founded his solution on the principle of the maximum, it will not now be difficult to produce an analytical formula of this sort which, when made equal to its maximum, yields the true value of x. Let us use for this purpose the form first discovered

$$r^2(x-a)+r^2(x-b)+r^2(x-c)+\text{etc.}$$
$$-(x-a)^3-(x-b)^3-(x-c)^3-\text{etc.}=0,$$

which may be regarded as the differential of some formula which is to be raised to its maximum. This formula itself will emerge, if this expression is put in the form of a differential and integrated. Multiplying by $4dx$ and integrating we obtain

$$2r^2(x-a)^2+2r^2(x-b)^2+2r^2(x-c)^2+\text{etc.}$$
$$-(x-a)^4-(x-b)^4-(x-c)^4-\text{etc.}+\text{constant.}$$

If we assume $-nr^4$ as the constant, there being n observations, by change of sign the following formula results

$$\{r^2-(x-a)^2\}^2+\{r^2-(x-b)^2\}^2+\{r^2-(x-c)^2\}^2+\text{etc.}$$

12. In place therefore of the formula which our distinguished friend Bernoulli thought should be made equal to its maximum we have now arrived at another formula very well suited to the nature of the question, which when brought to its maximum gives the true value of x, since this formula is obtained by adding together the squares of all the degrees of goodness.

13. To furnish an example of our method, let us consider the observations by which the longitude of the observatory of St Petersburg is deduced from the difference between the meridians of the observatories of Paris and St Petersburg. These are reported as follows:

I	1° 51′ 50″	IV	1° 51′ 50″
II	1° 51′ 52″	V	1° 51′ 50″
III	1° 51′ 39″	VI	1° 51′ 50″

Taking the arithmetic mean of these in the usual way we obtain 1° 51′ 48½″.

14. Now let us apply our formulae to this case, taking $\Pi = 1°\,51'\,48\tfrac{1}{2}''$. The values of our six letters a, b, c, d, e, f will be

$$a=1\tfrac{1}{2},\quad b=3\tfrac{1}{2},\quad c=-9\tfrac{1}{2},\quad d=1\tfrac{1}{2},\quad e=1\tfrac{1}{2},\quad f=1\tfrac{1}{2}.$$

Their sum $A = 0$; the sum of the squares B is found to be $\frac{1}{2}(223)$; the sum of the cubes $= -801$. Hence our equation for $n = 6$ will be

$$12ux^2 - 6u^2x + 801 + 334\tfrac{1}{2}x = 0.$$

15. Now let us *define* the number u from a case which the author of the observations thinks should be rejected, such as $1^\circ\,52'\,20''$, which gives $u = 31\frac{1}{2}$. Let us suppose that $u = 30$, making our quadratic equation

$$360x^2 - 5065\tfrac{1}{2}x + 801 = 0$$

instead of which we may write in round figures

$$36x^2 = 500x - 80.$$

From this
$$x = \frac{250 \pm \sqrt{59{,}620}}{36},$$

that is either
$$x = \frac{250+244}{36} = 14 \quad \text{or} \quad x = \frac{250-244}{36} = \frac{1}{6}.$$

The latter value only can be considered, and might have been obtained immediately by neglecting the first term in the equation: the value of x would then have been $\frac{50}{8}$ or approximately $\frac{1}{6}$. The required difference of the meridians will therefore be $1^\circ\,51'\,48\frac{2}{3}''$.

16. Again, suppose that observation had been rejected which gave $1^\circ\,51'\,0''$: we then have $u = -48\frac{1}{2}$. Let us take $u = -48$, giving the equation

$$-576x^2 - 13{,}489\tfrac{1}{2}x + 801 = 0.$$

Neglecting the first term we obtain $x = \frac{8}{1385} = \frac{1}{177}$. Now since this observation would have deserved to be rejected, if u had been in the neighbourhood of -300, hence, carrying out the calculation as before, x would have come out as about $\frac{1}{6}$. It is clear from this that in this case we could have been content with the common rule, since not even a second's difference is involved.

17. Since, however, the third of these observations differs so much from the others, it will perhaps be convenient to set the limit not far from it. If we were to do this for the case $1^\circ\,51'\,33\frac{1}{2}''$, $u = -15''$, our equation would accordingly be

$$-180x^2 - 1000x + 80 = 0,$$

the smaller root of which equation will be $\frac{12}{118} = \frac{2}{3}$. Hence the difference of the meridians would have come out as $1^\circ\,51'\,49\frac{1}{6}''$. It is once more clear from this case that no notable error is to be feared, unless we make a quite monstrous mistake in assuming a value for u. In this matter it will suffice to note that nu^2 must always be much larger than $3B$.

18. In particular this method deserves to be applied to those observations from which the learned Lexell not long ago determined the parallax of the sun. From these we take, purely by way of example, the following four conclusions drawn from the observations, namely (I) 8·52; (II) 8·43; (III) 8·86; (IV) 8·28. Taking the arithmetic mean of these we get 8·52. If therefore we put $\Pi = 8{\cdot}52$ the values of the four letters a, b, c, d can be fixed as follows:

$$a = 1, \quad b = 9, \quad c = -34, \quad d = +24$$

so that the sum comes out as $A = 0$.† All these numbers of course denote hundredths of a second. The sum of the squares $B = 1814$, the sum of the cubes $C = -24{,}750$.

† According to his original usage all these signs should be reversed. Presumably a has been rounded up to unity to make the sum of deviations zero.

19. If we now assume as the term where the degree of goodness vanishes $u = 40$, our equation emerges as

$$320x^2 - 948x + 24{,}750 = 0.$$

From this the value of x itself comes out as imaginary. Let us accordingly assume $u = 50$; the equation will then become

$$400x^2 - 10{,}000x + 24{,}750 + 5442x = 0$$

and we still arrive at an imaginary result. If, however, we take $u = 60$ the smaller value of x will be $3\frac{5}{12}$, which might seem to be too large. If we admit it the parallax of the sun would be 8·555. But let us note that larger values of u give smaller values for x. Since the application of this method is so vague, we may well doubt whether in this fashion we can arrive any closer to the truth, and perhaps it will suffice to have learnt at any rate, whether the value of x will come out positive or negative.

20. In this case, to be sure, we have seen that the value of x is certainly positive, since we have found a negative number for C. Hence we may profitably observe in general that whenever C comes out positive, x becomes negative, while if C is negative the value of x will be positive. In either case it must of necessity be so small that the result will hardly differ from the common rule. This at any rate can be added, that the larger the number C, the greater must necessarily be the value of x. For if the sum of the cubes C actually vanished, then x would always $= 0$, whatever value is accepted for u, just as the common rule requires.

21. Thus, notwithstanding the uncertainty produced by the number u, it seems that something reasonably probable can be laid down even if we cannot reach certainty, if we pay attention to the following points. First, it is certain that whenever the sum of the cubes $C = 0$, x will always $= 0$. Secondly, the larger the quantity C, the larger will be the value of x itself, with the opposite sign. Thirdly, it is clear enough that the quantity nu^2 must be very much greater than the quantity $3B$. In view of this we can lay it down with reasonable probability that $x = -C/\lambda nB$, where the number λ, it is true, is left to our judgement. However, it will meet all cases and depart hardly at all from the truth, if we put $\lambda = 2$ or at most $\lambda = 3$. The resulting difference will usually be so unimportant that we hardly need consider it. For the case where the greatest error is to be feared would undoubtedly be that in which several observations, i in number, agree entirely in each giving the value a, while the one remaining observation gives $-ia$, so that the sum of them all $A = 0$. The sum of the squares $B = ia^2 + i^2a^2 = i(i+1)a^2$; the sum of the cubes $= ia^3 - i^3a^3 = -i(i^2-1)a^3$. For $n = i+1$ our formula gives

$$x = +\frac{i(i^2-1)a}{\lambda i(i+1)^2} = \frac{(i-1)a}{\lambda(i+1)}.$$

If therefore i is very large and we take $\lambda = 2$ the result is $x = \frac{1}{2}a$. In the earlier example where $n = b$, $B = 11\frac{1}{2}$ and $C = -801$, $x = +801/(12 \times 11\frac{1}{2}) = \frac{3}{5}$ approximately. In the second, where $n = 4$, $B = 1814$, $C = -24{,}750$, $x = 24{,}750/(8 \times 1814) = \frac{8}{5}$ approximately. These values do not appear to involve anything absurd.

If, however, anyone thinks that it would be more reasonable to take $\lambda = 3$, I hardly think the difference is worth arguing about, since the very nature of the observations does not admit of a greater degree of precision.

Biometrika (1961), **48**, 1 *and* 2, *p.* 19
Printed in Great Britain

The variance of Spearman's rho in normal samples

BY F. N. DAVID† AND C. L. MALLOWS‡
University College London

Kendall (1949) and David, Kendall & Stuart (1951), modified by Fieller, Hartley & Pearson (1957), first gave an approximation to the variance of Spearman's rho, r_s, in samples of n from a normal bivariate population in which the product moment correlation is ρ. This modified value is

$$\sigma^2_{r_s} \doteqdot \frac{1}{n-1}[1-1{\cdot}563{,}465\rho^2+0{\cdot}304{,}743\rho^4+0{\cdot}155{,}286\rho^6 + 0{\cdot}061{,}552\rho^8+0{\cdot}022{,}099\rho^{10}+0{\cdot}019{,}785\rho^{12}].$$

Recently we have had occasion to require a closer approximation to this variance and have obtained one which is exact for n and to order ρ^{12} for ρ. The leading terms of our expression agree with those of Kendall's. Many of the actual quantities calculated to achieve our final result will be useful in other aspects of the rank problem. Accordingly we give these auxiliary quantities in Appendices so that they may be used for other purposes.

Assume n pairs of observations $\{x_i, y_i\}$ randomly and independently drawn from a bivariate normal population in which the product moment correlation is ρ. Let X_i be the rank of x_i and Y_i the rank of y_i. Using the technique also applied by Moran (1948) for finding $\mathscr{E}(r_s)$, we define a function $H(t)$ such that

$$H(t) = 0 \quad (t \leqslant 0)$$
$$= 1 \quad (t > 0).$$

Accordingly we may write

$$X_i - 1 = \sum_{j=1}^{n} H(x_i - x_j), \quad Y_i - 1 = \sum_{m=1}^{n} H(y_i - y_m)$$

with

$$s = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} H(x_i - x_j)\,H(y_i - y_m),$$

$$r_s = \frac{s - \{\frac{1}{2}(n-1)\}^2}{n}\left(\frac{12}{n^2-1}\right).$$

Since $\mathscr{E}(s)$ is known the variance of r_s depends on the evaluation of $\mathscr{E}(s^2)$. This involves expanding the square of the triple sum, s, taking expectations term by term and collecting together like terms. A list of the different terms involved and the numbers of each can easily be compiled using David & Kendall's symmetric function tables (1949) or the new bivariate symmetric function tables recently compiled by Barton, David & Fix. Each term is the expectation of a product of four H's. We require the probability that each of these products is positive, given that the $\{x_i y_i\}$ are randomly and independently drawn, which means simply the evaluation of the positive quadrant of the quadrivariate normal integral in each case. This is not as formidable as it sounds since many of the products have the same variance-covariance matrix and in the event only 23 evaluations were required.

† With the partial support of the Office of Naval Research while at the University of California, Berkeley.

‡ Now with the Bell Telephone Company, Murray Hill, New Jersey.

Moran (1948) and Kendall (1941) have both given series expansions whereby the positive quadrant of the quadrivariate normal integral may be evaluated. These series are however, for ρ at all large, very slow to converge, and this makes them unsatisfactory for present purposes. Here we make use of a procedure put forward by Plackett (1954). Let the correlation matrix be

$$R = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1 & \rho_{23} & \rho_{24} \\ & & 1 & \rho_{34} \\ & & & 1 \end{pmatrix}$$

and its inverse

$$C = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ & C_{22} & C_{23} & C_{24} \\ & & C_{33} & C_{34} \\ & & & C_{44} \end{pmatrix}.$$

Write

$$X_{pq} = \frac{C_{pq}}{\sqrt{(C_{pp}C_{qq})}}$$

to form the matrix

$$X = \begin{pmatrix} 1 & X_{12} & X_{13} & X_{14} \\ & 1 & X_{23} & X_{24} \\ & & 1 & X_{34} \\ & & & 1 \end{pmatrix}.$$

Now each of the elements of R will be, for this particular problem, either a constant or ρ with a constant multiplier. Let K be the matrix R when ρ is put equal to zero, and L be the matrix when ρ is put equal to unity. Then, writing Φ for the integral over the positive quadrant,

$$\Phi(R) = \Phi(K) + \frac{1}{4\pi^2} \sum_{i<j} \int_0^\rho \frac{L_{ij}\, d\rho}{\sqrt{(1 - L_{ij}^2 \rho^2)}} \cos^{-1}(X_{pq}),$$

where p, q are the elements of the permutation 1, 2, 3, 4 which are *not* i, j and L_{ij} are those elements of L which are different from those of K. This procedure follows directly from Plackett's equations (6) or (7). It enables an expansion to be made about points other than $\rho_{ij} = 0$, and in most cases gives a quickly convergent series.

Example. Sundrum (1953) has given the distribution of Spearman's rho for $n = 3$ by an argument which he does not set out. His results can be shown to follow easily from an application of the method of the previous section. Assume three sets of bivariate normal variables $(x_i y_i)$, $i = 1, 2, 3$, with correlation ρ. Let $(x_1 y_2, x_2 y_3, x_3 y_1)$ correspond to a ranking (12, 23, 31) so that the sample Spearman's rho r_s is $-\frac{1}{2}$. Then

$$P\{r_s = -\tfrac{1}{2}\} = P\{(x_1 - x_2)(x_2 - x_3)(y_2 - y_3)(y_3 - y_1) > 0\}.$$

Writing

$$X_1 = x_1 - x_2, \quad X_2 = x_2 - x_3, \quad Y_1 = y_2 - y_3, \quad Y_2 = y_3 - y_1$$

the variance covariance matrix of $X_1 X_2 Y_1 Y_2$ and the correlation matrix is

$$\begin{pmatrix} 2 & -1 & -\rho & -\rho \\ & 2 & 2\rho & -\rho \\ & & 2 & -1 \\ & & & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2}\rho & -\frac{1}{2}\rho \\ & 1 & \rho & -\frac{1}{2}\rho \\ & & 1 & -\frac{1}{2} \\ & & & 1 \end{pmatrix} = R.$$

The inverse is

$$C = \begin{pmatrix} \frac{4}{3(1-\rho^2)} & \frac{2}{3(1-\rho^2)} & \frac{2\rho}{3(1-\rho^2)} & \frac{4\rho}{3(1-\rho^2)} \\ & \frac{4}{3(1-\rho^2)} & \frac{-2\rho}{3(1-\rho^2)} & \frac{2\rho}{3(1-\rho^2)} \\ & & \frac{4}{3(1-\rho^2)} & \frac{2}{3(1-\rho^2)} \\ & & & \frac{4}{3(1-\rho^2)} \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2}\rho & \rho \\ & 1 & -\frac{1}{2}\rho & \frac{1}{2}\rho \\ & & 1 & \frac{1}{2} \\ & & & 1 \end{pmatrix}.$$

The K and L matrices are

$$K = \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ & 1 & 0 & 0 \\ & & 1 & -\frac{1}{2} \\ & & & 1 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ & 1 & 1 & -\frac{1}{2} \\ & & 1 & -\frac{1}{2} \\ & & & 1 \end{pmatrix},$$

whence $$\Phi(K) = \tfrac{1}{36}$$

and

$$\Phi(R) = \frac{1}{36} + \frac{1}{4\pi^2}\left\{-2\int_0^\rho \frac{\frac{1}{2}}{\sqrt{(1-\frac{1}{4}\rho^2)}}\cos^{-1}(\tfrac{1}{2}\rho)\,d\rho - \int_0^\rho \frac{\frac{1}{2}}{\sqrt{(1-\frac{1}{4}\rho^2)}}\cos^{-1}(-\tfrac{1}{2}\rho)\,d\rho \right.$$
$$\left. + \int_0^\rho \frac{1}{\sqrt{(1-\rho^2)}}\cos^{-1}(\rho)\,d\rho\right\}$$
$$= \frac{1}{36} + \frac{1}{8\pi}[-3\sin^{-1}\tfrac{1}{2}\rho + \sin^{-1}\rho] + \frac{1}{8\pi^2}[(\sin^{-1}\tfrac{1}{2}\rho)^2 - (\sin^{-1}\rho)^2].$$

The probability of the ranking (12, 23, 31) or of any permutation of the 3 pairs, is $6\Phi(R)$ which agrees with Sundrum.

Essentially the same technique is used to find the probability of each of the products in Appendix 1. Each term consists of the product of four normal variables and the probability that this product is positive is required. There are fifty-eight of these products but there are many common correlation matrices and the work reduces to the evaluation of twenty-three such. Of these twenty-three probabilities some can be written down at sight, others may be evaluated exactly, while twelve of them involve the evaluation of complicated trigonometric functions. They are not elliptic integrals in the strict sense. The correlation matrices and the associated probability in the positive quadrant are set out in Appendix 2. It will be noticed that eleven integrals have no explicit solution. We have given series expansions for these integrals in Appendix 3 and writing

$$S_1 = \sin^{-1}\rho, \quad S_2 = \sin^{-1}\tfrac{1}{2}\rho$$

will use the notation $A(S_1)$, $A(S_2)$, etc., to indicate that the series is either in powers of $\sin^{-1}\rho$ or $\sin^{-1}\frac{1}{2}\rho$.

Collecting the terms in Appendix 1 with the help of Appendices 2 and 3 we have

$$\mathscr{E}(s^2) = \frac{n^2(n-1)}{144}(9n^3-26n^2+29n-8)+\frac{n^{(2)}n(n-1)^2}{4\pi}\{S_2(n-2)+S_1\}$$
$$+\frac{1}{4\pi^2}\{n^{(6)}S_2^2+2n^{(5)}S_1S_2+n^{(4)}S_1^2+n^{(3)}(6S_1^2-6S_2^2)\}$$
$$+\frac{1}{\pi_2}[n^{(5)}\{2A(S_2)+C(S_2)+J(S_2)+\tfrac{1}{2}K(S_2)\}$$
$$+n^{(4)}\{2G(S_1)+D(S_1)-2H(S_2)+2I(S_2)-2E(S_2)+3F(S_2)+B(S_2)\}].$$

Remembering that $$\mathscr{E}(r_s) = \frac{6}{\pi(n+1)}[(n-2)S_2+S_1]$$

we have, after reduction

$$\text{var}\,(r_s) = \frac{1}{n-1}-\frac{36}{\pi^2n(n-1)(n+1)^2}[3(n-2)(3n^2-15n+22)S_2^2$$
$$+12(n-2)^2S_1S_2-2(n-3)S_1^2]$$
$$+\frac{144(n-2)(n-3)}{\pi^2n(n-1)(n+1)^2}[(n-4)\{2A(S_2)+C(S_2)+J(S_2)+\tfrac{1}{2}K(S_2)\}$$
$$+\{2G(S_1)+D(S_1)-2H(S_2)+2I(S_2)-2E(S_2)+3F(S_2)+B(S_2)\}].$$

For purposes of computation this may be written as

$$\text{var}\,(r_s) = \frac{1}{n-1}-\frac{36}{\pi^2n(n-1)(n+1)^2}[3(n-2)(3n^2-15n+22)S_2^2$$
$$+12(n-2)^2S_1S_2-2(n-3)S_1^2]$$
$$+\frac{144(n-2)(n-3)}{\pi^2n(n-1)(n+1)^2}[(n-4)\alpha+\beta+\gamma], \qquad (X)$$

where

$$\alpha = 1{\cdot}821,3672S_2^2+0{\cdot}477,0655^{-}S_2^4+0{\cdot}439,3170S_2^6+0{\cdot}467,0222S_2^8$$
$$+0{\cdot}565,8339S_2^{10}+0{\cdot}759,3293S_2^{12},$$

$$\beta = 3{\cdot}821,3672S_2^2+2{\cdot}682,0510S_2^4+3{\cdot}395,6439S_2^6+6{\cdot}662,0844S_2^8$$
$$+10{\cdot}522,7533S_2^{10}+22{\cdot}784,3043S_2^{12},$$

$$\gamma = 0{\cdot}744,0169S_1^2-0{\cdot}044,4207S_1^4-0{\cdot}002,1383S_1^6+0{\cdot}000,1221S_1^8$$
$$+0{\cdot}000,0407S_1^{10}+0{\cdot}000,0021S_1^{12}.$$

The coefficients in these series are known exactly and may be calculated to as many decimal places as desired.

When $\rho = 0$ it is seen that (X) gives the value $1/(n-1)$ for var (r_s) as indeed it must. When $|\rho| = 1$ the value of var (r_s) is not zero, as it should be, but $-0{\cdot}003,320/(n-1)$ a nonsensical value arising from the fact that the convergence of the series β is too slow to permit any reasonable approximation. The range over which (X) will be useful will depend on both n and ρ as well as on the degree of accuracy required. Arithmetic investigations suggest that a suitable rule for n of reasonable size and for five figure accuracy in var (r_s), $|\rho|$ should not

exceed 0·8. Specimen calculations for $|\rho| = 0{\cdot}5$ and 0·6 will serve for illustration. Since all the terms are exact except for α, β and γ we give these only.

$|\rho| = 0{\cdot}5$,

$$\alpha = 0{\cdot}116{,}289 + 0{\cdot}001{,}945^{-} + 0{\cdot}000{,}114 + 0{\cdot}000{,}008 + 0{\cdot}000{,}001 + 0{\cdot}000{,}000 = 0{\cdot}118{,}357,$$
$$\beta = 0{\cdot}243{,}984 + 0{\cdot}010{,}933 + 0{\cdot}000{,}884 + 0{\cdot}000{,}111 + 0{\cdot}000{,}011 + 0{\cdot}000{,}001^{(5)} = 0{\cdot}255{,}925,$$
$$\gamma = 0{\cdot}203{,}977 - 0{\cdot}003{,}339 - 0{\cdot}000{,}044 + 0{\cdot}000{,}001 = 0{\cdot}200{,}594;$$

$|\rho| = 0{\cdot}6$,

$$\alpha = 0{\cdot}169{,}091 + 0{\cdot}004{,}112 + 0{\cdot}000{,}352 + 0{\cdot}000{,}035^{-} + 0{\cdot}000{,}004 + 0{\cdot}000{,}001 = 0{\cdot}173{,}594,$$
$$\beta = 0{\cdot}354{,}767 + 0{\cdot}023{,}116 + 0{\cdot}002{,}717 + 0{\cdot}000{,}494 + 0{\cdot}000{,}073 + 0{\cdot}000{,}015^{-} = 0{\cdot}381{,}182,$$
$$\gamma = 0{\cdot}308{,}093 - 0{\cdot}007{,}617 - 0{\cdot}000{,}152 + 0{\cdot}000{,}004 = 0{\cdot}300{,}328.$$

Thus, writing $N = n-1$ for conciseness, we have, for example, when $|\rho| = 0{\cdot}5$, the alternative forms

$$\operatorname{var}(r_s) = \frac{1}{N} - \frac{36}{\pi^2(N+1)N(N+2)^2}[N^3(0{\cdot}101{,}198) + N^2(0{\cdot}303{,}626) + N(0{\cdot}186{,}232) - 0{\cdot}042{,}742]$$
$$= \frac{1}{N} - \frac{1}{N(N+1)(N+2)^2}[N^3(0{\cdot}369{,}13) + N^2(1{\cdot}107{,}49) + N(0{\cdot}679{,}34) - 0{\cdot}155{,}90]$$
$$= \frac{0{\cdot}630{,}87}{N} + \frac{0{\cdot}738{,}16}{N^2} - \frac{1{\cdot}417{,}10}{N^3} + \frac{2{\cdot}812{,}64}{N^4}. \qquad (Y)$$

(Y) would seem to be the appropriate formula for the calculation of $\operatorname{var}(r_s;\ |\rho| = 0{\cdot}5)$ for any $N = n-1$.

We have given some thought to the possibility of alternate expansions of the integrals A, B, etc., when ρ is large. When $|\rho| = 1$, $S_1 = \frac{1}{2}\pi$, $S_2 = \frac{1}{6}\pi$ and

$$\operatorname{var}(r_s) = 0 = \frac{1}{n-1} - \frac{36}{\pi^2 n(n-1)(n+1)^2}[3(n-2)(3n^2-15n+22)\tfrac{1}{36}\pi^2 + (n-2)^2\pi^2 - \tfrac{1}{2}(n-3)\pi^2]$$
$$+ \frac{144}{\pi^2 n(n-1)(n+1)^2}[(n-4)\{2A(\tfrac{1}{6}\pi) + C(\tfrac{1}{6}\pi) + J(\tfrac{1}{6}\pi) + \tfrac{1}{2}K(\tfrac{1}{2}\pi)\}$$
$$+ \{2G(\tfrac{1}{2}\pi) + D(\tfrac{1}{2}\pi) - 2H(\tfrac{1}{6}\pi) + 2I(\tfrac{1}{6}\pi) - 2E(\tfrac{1}{6}\pi) + 3F(\tfrac{1}{6}\pi) + B(\tfrac{1}{6}\pi)\}].$$

Let
$$A^*(S_2) = \int_0^{\frac{1}{6}\pi} \sin^{-1}\left(\frac{\sin 2x}{\sqrt{(1+2\cos 2x)}}\right)dx - \int_0^{\sin^{-1}\frac{1}{2}\rho} \sin^{-1}\left(\frac{\sin 2x}{\sqrt{(1+2\cos 2x)}}\right)dx$$
$$= \int_{\sin^{-1}\frac{1}{2}\rho}^{\frac{1}{6}\pi} \sin^{-1}\left(\frac{\sin 2x}{\sqrt{(1+2\cos 2x)}}\right)dx$$

with similar meanings for B, C, etc. Then

$$\operatorname{var}(r_s) = \frac{144}{\pi^2 n(n-1)(n+1)^2}[(n-4)\{2A^*(S_2) + C^*(S_2) + J^*(S_2) + \tfrac{1}{2}K^*(S_2)\}$$
$$+ \{2G^*(S_1) + D^*(S_1) - 2H^*(S_2) + 2I^*(S_2) - 2E^*(S_2) + 3F^*(S_2) + B^*(S_2)\}]$$
$$- \frac{36}{\pi^2 n(n-1)(n+1)^2}[3(n-2)(3n^2-15n+22)(\tfrac{1}{36}\pi^2 - S_2^2)$$
$$+ (n-2)^2(\pi^2 - 12S_1S_2) - \tfrac{1}{2}(n-3)(\pi^2 - 4S_1^2)].$$

Some simplification is possible, as indeed it also is in (X), by adding integrals together, but this only reduces the number of series to be evaluated by one or two. We expand about the

upper limit and obtain series for A^*, B^*, C^*, E^*, F^*, H^*, I^*, J^*, K^* in terms of powers of $(\frac{1}{6}\pi - \sin^{-1}\frac{1}{2}\rho)$ and for G^*, D^* in powers of $(\frac{1}{2}\pi - \sin^{-1}\rho)$. The arithmetic involved in the algebraic expansions of this kind is, however, lengthy and we do not propose to finish this part of our investigation until the necessity for the variance of r_s for ρ large becomes apparent.

One further way of expressing var (r_s) might be mentioned. The quantities A, B, C, etc., may be expressed as a power series in ρ instead of in S_1 and S_2. If we do this we have

$$\begin{aligned}\text{var}\,(r_s) = \frac{1}{n-1} - \frac{36}{\pi^2 n(n-1)(n+1)^2}&[3(n-2)(3n^2-15n+22)S_2^2 \\ &+ 12(n-2)^2 S_1 S_2 - 2(n-3)S_1^2] \\ + \frac{144(n-2)(n-3)}{\pi^2 n(n-1)(n+1)^2}&[n(0{\cdot}455{,}342\rho^2 + 0{\cdot}067{,}762\rho^4 + 0{\cdot}016{,}893\rho^6 + 0{\cdot}005{,}223\rho^8 \\ &+ 0{\cdot}001{,}839\rho^{10} + 0{\cdot}000{,}710\rho^{12}) + (-0{\cdot}122{,}008\rho^2 \\ &+ 0{\cdot}179{,}778\rho^4 + 0{\cdot}124{,}555\rho^6 + 0{\cdot}087{,}275\rho^8 \\ &+ 0{\cdot}059{,}080\rho^{10} + 0{\cdot}042{,}552\rho^{12})].\end{aligned}$$

If S_2 and S_1 are also expanded as series in ρ then

$$\begin{aligned}\text{var}\,(r_s) = \frac{1}{n-1} + \frac{36}{\pi^2 n(n-1)(n+1)^2}&[n^3(-0{\cdot}428{,}632{,}79\rho^2 + 0{\cdot}083{,}546{,}97\rho^4 + 0{\cdot}042{,}572{,}46\rho^6 \\ &+ 0{\cdot}016{,}874{,}74\rho^8 + 0{\cdot}006{,}640{,}71\rho^{10} + 0{\cdot}002{,}706{,}55\rho^{12}) \\ &+ n^2(0{\cdot}155{,}130{,}10\rho^2 - 0{\cdot}0573{,}622{,}93\rho^4 - 0{\cdot}184{,}434{,}07\rho^6 \\ &- 0{\cdot}022{,}717{,}32\rho^8 + 0{\cdot}007{,}575{,}24\rho^{10} + 0{\cdot}013{,}298{,}83\rho^{12}) \\ &+ n(0{\cdot}368{,}372{,}59\rho^2 + 0{\cdot}447{,}388{,}82\rho^4 - 0{\cdot}084{,}275{,}74\rho^6 \\ &- 0{\cdot}279{,}299{,}01\rho^8 - 0{\cdot}199{,}433{,}75\rho^{10} - 0{\cdot}138{,}610{,}60\rho^{12}) \\ &+ (0{\cdot}071{,}796{,}77\rho^2 + 0{\cdot}064{,}671{,}62\rho^4 + 0{\cdot}210{,}152{,}57\rho^6 \\ &+ 0{\cdot}285{,}897{,}98\rho^8 + 0{\cdot}317{,}044{,}25\rho^{10} + 0{\cdot}079{,}237{,}33\rho^{12})].\end{aligned} \tag{Z}$$

The terms in $1/n$ in this series agree with those already given by David *et al.* (1951), with the exception of the coefficient of ρ^{10}, which we find to be 0·0242,224 against their 0·0220,992.

In the course of this work we have investigated many methods which have been proposed for the evaluation of the positive quadrant of the quadrivariate normal integral but have found none of greater utility than that proposed by Plackett. It is clear that his given technique can be similarly modified for the evaluation of the six-dimensional normal integral, in this case two integrals being required for the evaluation of each set of concordances. We are at present investigating this process with the idea of obtaining the p.d.f. of r_s for the case $n = 4$. Barbara Snow, who checked all the algebra in the present paper is hoping to apply the same method to obtain the third moment of Kendall's 'tau' coefficient, the enumeration for which has been done by Sundrum (1953), and the third moment of Spearman's 'rho'. This latter will, however, take some time to accomplish although it is not difficult in principle.

We would like to acknowledge help given in checking algebra by Barbara Snow and Evelyn Fix. D. E. Barton contributed many ideas for the expansions in terms of $(\frac{1}{6}\pi - \sin^{-1}\frac{1}{2}\rho)$ which we have not pursued to completion.

References

David, F. N. & Kendall, M. G. (1949). Tables of symmetric functions. Part I. *Biometrika*, **36**, 431–49.

David, S. T., Kendall, M. G. & Stuart, A. (1951). Some questions of distribution in the theory of rank correlation. *Biometrika*, **38**, 131–40.

Fieller, E. C., Hartley, H. O. & Pearson, E. S. (1957). Tests for rank correlation coefficients. I. *Biometrika*, **44**, 470–81.

Kendall, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalization. *Biometrika*, **32**, 196–8.

Kendall, M. G. (1949). Rank and product moment correlation. *Biometrika*, **36**, 177–93.

Moran, P. A. P. (1948). Rank correlation and product moment correlation. *Biometrika*, **35**, 203–6.

Plackett, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351–60.

Sundrum, R. M. (1953). Moments of the rank correlation coefficient tau in the general case. *Biometrika*, **49**, 409–20.

Appendix 1

[The numbers 1, 2, 3, 4, 5, 6, are used merely to differentiate the x's and y's.]

Correlation matrix	No. of terms	Type	Partition
a	$n^{(6)}$	$(x_1-x_2)(y_1-y_3)(x_4-x_5)(y_4-y_6)$	(1^6)
b	$n^{(5)}$	$(x_1-x_2)(y_1-y_2)(x_3-x_4)(y_3-y_5)$	(2^11^4)
c		$(x_1-x_2)(y_1-y_3)(x_1-x_4)(y_1-y_5)$	
d		$(x_1-x_2)(y_1-y_3)(x_2-x_4)(y_2-y_5)$	
d		$(x_1-x_2)(y_1-y_3)(x_3-x_4)(y_3-y_5)$	
d		$(x_1-x_2)(y_1-y_3)(x_4-x_1)(y_4-y_5)$	
e		$(x_1-x_2)(y_1-y_3)(x_4-x_2)(y_4-y_5)$	
f		$(x_1-x_2)(y_1-y_3)(x_4-x_3)(y_4-y_5)$	
d		$(x_1-x_2)(y_1-y_3)(x_4-x_5)(y_4-y_1)$	
f		$(x_1-x_2)(y_1-y_3)(x_4-x_5)(y_4-y_2)$	
e		$(x_1-x_2)(y_1-y_3)(x_4-x_5)(y_4-y_3)$	
b		$(x_1-x_2)(y_1-y_3)(x_4-x_5)(y_4-y_5)$	
g	$n^{(4)}$	$(x_1-x_2)(y_1-y_3)(x_1-x_3)(y_1-y_4)$	(2^21^2)
h		$(x_1-x_2)(y_1-y_2)(x_3-x_4)(y_3-y_1)$	
i		$(x_1-x_2)(y_1-y_3)(x_3-x_4)(y_3-y_2)$	
j		$(x_1-x_2)(y_1-y_2)(x_3-x_4)(y_3-y_4)$	
g		$(x_1-x_2)(y_1-y_3)(x_1-x_4)(y_1-y_4)$	
h		$(x_1-x_2)(y_1-y_3)(x_2-x_4)(y_2-y_4)$	
h		$(x_1-x_2)(y_1-y_3)(x_3-x_4)(y_3-y_4)$	
g		$(x_1-x_2)(y_1-y_3)(x_1-x_4)(y_1-y_2)$	
k		$(x_1-x_2)(y_1-y_3)(x_1-x_2)(y_1-y_4)$	
g		$(x_1-x_2)(y_1-y_3)(x_1-x_3)(y_1-y_4)$	
i		$(x_1-x_2)(y_1-y_3)(x_2-x_3)(y_2-y_4)$	
i		$(x_1-x_2)(y_1-y_3)(x_4-x_1)(y_4-y_2)$	
l		$(x_1-x_2)(y_1-y_3)(x_4-x_1)(y_4-y_3)$	
m		$(x_1-x_2)(y_1-y_3)(x_4-x_2)(y_4-y_3)$	
n		$(x_1-x_2)(y_1-y_3)(x_2-x_4)(y_2-y_1)$	
k		$(x_1-x_2)(y_1-y_3)(x_1-x_4)(y_1-y_3)$	
l		$(x_1-x_2)(y_1-y_3)(x_2-x_4)(y_2-y_3)$	
h		$(x_1-x_2)(y_1-y_3)(x_3-x_1)(y_3-y_4)$	
n		$(x_1-x_2)(y_1-y_3)(x_3-x_1)(y_3-y_4)$	
l		$(x_1-x_2)(y_1-y_3)(x_2-x_3)(y_3-y_4)$	
l		$(x_1-x_2)(y_1-y_3)(x_4-x_2)(y_4-y_1)$	
i		$(x_1-x_2)(y_1-y_3)(x_4-x_3)(y_4-y_1)$	
o		$(x_1-x_2)(y_1-y_3)(x_4-x_3)(y_4-y_2)$	
p	$n^{(4)}$	$(x_1-x_2)(y_1-y_2)(x_2-x_3)(y_2-y_4)$	(31^3)
q		$(x_1-x_2)(y_1-y_3)(x_3-x_2)(y_3-y_4)$	
q		$(x_1-x_2)(y_1-y_2)(x_3-x_4)(y_3-y_2)$	
p		$(x_1-x_2)(y_1-y_3)(x_4-x_1)(y_4-y_1)$	
q		$(x_1-x_2)(y_1-y_3)(x_4-x_2)(y_4-y_2)$	
q		$(x_1-x_2)(y_1-y_3)(x_4-x_3)(y_4-y_3)$	
r	$n^{(3)}$	$(x_1-x_2)(y_1-y_2)(x_1-x_3)(y_1-y_3)$	(2^3)
s		$(x_1-x_2)(y_1-y_3)(x_1-x_2)(y_1-y_3)$	
r		$(x_1-x_2)(y_1-y_3)(x_1-x_3)(y_1-y_2)$	
t		$(x_1-x_2)(y_1-y_3)(x_3-x_1)(y_3-y_2)$	
t		$(x_1-x_2)(y_1-y_3)(x_2-x_3)(y_2-y_1)$	
u	$n^{(3)}$	$(x_1-x_2)(y_1-y_2)(x_1-x_2)(y_1-y_3)$	(321)
u		$(x_1-x_2)(y_1-y_2)(x_1-x_3)(y_1-y_2)$	
v		$(x_1-x_2)(y_1-y_2)(x_3-x_1)(y_3-y_2)$	
v		$(x_1-x_2)(y_1-y_2)(x_3-x_2)(y_3-y_1)$	
w		$(x_1-x_2)(y_1-y_2)(x_3-x_1)(y_3-y_1)$	
u		$(x_1-x_2)(y_1-y_3)(x_1-x_2)(y_1-y_2)$	
u		$(x_1-x_2)(y_1-y_3)(x_1-x_3)(y_1-y_3)$	
v		$(x_1-x_2)(y_1-y_3)(x_3-x_2)(y_3-y_2)$	
v		$(x_1-x_2)(y_1-y_3)(x_2-x_3)(y_2-y_3)$	
w		$(x_1-x_2)(y_1-y_2)(x_2-x_3)(y_2-y_3)$	
r	$n^{(3)}$	$(x_1-x_2)(y_1-y_2)(x_3-x_2)(y_3-y_2)$	(41^2)
x	$n^{(2)}$	$(x_1-x_2)(y_1-y_2)(x_1-x_2)(y_1-y_2)$	(42)

APPENDIX 2

	Correlation matrix	Area in positive quadrant of quadrivariate normal surface
a	$\begin{pmatrix} \frac{1}{2}\rho & 0 & 0 \\ & 0 & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\left(\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\frac{\rho}{2}\right)^2$
b	$\begin{pmatrix} \rho & 0 & 0 \\ & 0 & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\left(\frac{1}{4}+\frac{\sin^{-1}\frac{1}{2}\rho}{2\pi}\right)\left(\frac{1}{4}+\frac{\sin^{-1}\rho}{2\pi}\right)$
c	$\begin{pmatrix} \frac{1}{2}\rho & \frac{1}{2} & \frac{1}{2}\rho \\ & \frac{1}{2}\rho & \frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{9}+\frac{1}{2\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{\pi^2}\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{\sin x}{1+2\cos 2x}\right)dx$
d	$\begin{pmatrix} \frac{1}{2}\rho & -\frac{1}{2} & -\frac{1}{2}\rho \\ & 0 & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{24}+\frac{1}{8\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{4\pi^2}\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{\sin 2x}{\sqrt{(1+2\cos 2x)}}\right)dx$
e	$\begin{pmatrix} \frac{1}{2}\rho & \frac{1}{2} & 0 \\ & 0 & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{12}+\frac{1}{4\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{2\pi^2}\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{\sin 2x}{\sqrt{(1+2\cos 2x)}}\right)dx$
f	$\begin{pmatrix} \frac{1}{2}\rho & 0 & 0 \\ & \frac{1}{2}\rho & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{16}+\frac{3}{8\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{4\pi^2}\left\{2\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{\sin 2x}{2\sqrt{(\cos 2x)}}\right)dx + \int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{3\sin x-\sin 3x}{4\cos 2x}\right)dx\right\}$
g	$\begin{pmatrix} \rho & \frac{1}{2} & \frac{1}{2}\rho \\ & \frac{1}{2}\rho & \frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{9}+\frac{1}{8\pi}(\sin^{-1}\rho+3\sin^{-1}\frac{1}{2}\rho)+\frac{1}{4\pi^2}\left\{\int_0^{\sin^{-1}\rho}\sin^{-1}\left(\frac{\sin x}{3}\right)dx - 2\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x-1}{3(2\cos 2x+1)}}\right)dx + \int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{3+2\cos 2x}{1+2\cos 2x}\right)dx\right\}$
h	$\begin{pmatrix} \rho & 0 & -\frac{1}{2}\rho \\ & 0 & -\frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{24}+\frac{\sin^{-1}\rho}{8\pi}+\frac{1}{4\pi^2}\left\{\int_0^{\sin^{-1}\rho}\sin^{-1}\left(\frac{\sin x}{\sqrt{3}}\right)dx - \int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x-1}{3\cos 2x}}\right)dx + \int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x+1}{\cos 2x}}\right)dx\right\}$
i	$\begin{pmatrix} \frac{1}{2}\rho & 0 & \frac{1}{2}\rho \\ & -\frac{1}{2}\rho & -\frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{24}+\frac{1}{4\pi}\sin^{-1}\frac{\rho}{2}$
j	$\begin{pmatrix} \rho & 0 & 0 \\ & 0 & 0 \\ & & \rho \end{pmatrix}$	$\left(\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\rho\right)^2$
k	$\begin{pmatrix} \frac{1}{2}\rho & \frac{1}{2}\rho \\ & \frac{1}{2} \end{pmatrix}$	$\frac{1}{6}+\frac{1}{2\pi}\sin^{-1}\frac{\rho}{2}$
l	$\begin{pmatrix} \frac{1}{2}\rho & -\frac{1}{2} & 0 \\ & -\frac{1}{2}\rho & \frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{18}+\frac{1}{8\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{4\pi^2}\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{3+2\cos 2x}{1+2\cos 2x}\sin x\right)dx$

APPENDIX 2 (*cont.*)

	Correlation matrix	Area in positive quadrant of quadrivariate normal surface
m	$\begin{pmatrix} \frac{1}{2}\rho & \frac{1}{2} & 0 \\ & 0 & \frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{9}+\frac{1}{4\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{2\pi^2}\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\frac{3+2\cos 2x}{1+2\cos 2x}\right)dx$
n	$\begin{pmatrix} \frac{1}{2}\rho & -\frac{1}{2} & -\rho \\ & 0 & -\frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{36}+\frac{1}{8\pi}(2\sin^{-1}\frac{1}{2}\rho-\sin^{-1}\rho)+\frac{1}{4\pi^2}\left\{2\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\times\left(\sin x\sqrt{\frac{2\cos 2x-1}{3(2\cos 2x+1)}}\right)dx-\int_0^{\sin^{-1}\rho}\sin^{-1}\left(\frac{\sin x}{3}\right)dx\right\}$
o	$\begin{pmatrix} \frac{1}{2}\rho & 0 & \frac{1}{2}\rho \\ & \frac{1}{2}\rho & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{16}+\frac{1}{2\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{\pi^2}\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\frac{\sin x}{\cos 2x}\right)dx$
p	$\begin{pmatrix} \rho & -\frac{1}{2} & -\frac{1}{2}\rho \\ & -\frac{1}{2}\rho & -\frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{36}+\frac{1}{8\pi}(\sin^{-1}\rho-\sin^{-1}\frac{1}{2}\rho)+\frac{1}{4\pi^2}\left\{\int_0^{\sin^{-1}\rho}\sin^{-1}\left(\frac{\sin x}{3}\right)dx-2\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x-1}{3(2\cos 2x+1)}}\right)dx+\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\frac{(3+2\cos 2x)}{(1+2\cos 2x)}\right)dx\right\}$
q	$\begin{pmatrix} \rho & \frac{1}{2} & 0 \\ & \frac{1}{2}\rho & 0 \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{12}+\frac{1}{8\pi}(\sin^{-1}\rho+2\sin^{-1}\frac{1}{2}\rho)+\frac{1}{4\pi^2}\left\{\int_0^{\sin^{-1}\rho}\sin^{-1}\left(\frac{\sin x}{\sqrt{3}}\right)dx-\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x-1}{3\cos 2x}}\right)dx+\int_0^{\sin^{-1}\frac{1}{2}\rho}\sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x+1}{\cos 2x}}\right)dx\right\}$
r	$\begin{pmatrix} \rho & \frac{1}{2} & \frac{1}{2}\rho \\ & \frac{1}{2}\rho & \frac{1}{2} \\ & & \rho \end{pmatrix}$	$\frac{1}{9}+\frac{1}{4\pi}(\sin^{-1}\rho+\sin^{-1}\frac{1}{2}\rho)+\frac{1}{4\pi^2}\{(\sin^{-1}\rho)^2-(\sin^{-1}\frac{1}{2}\rho)^2\}$
s	$(\frac{1}{2}\rho)$	$\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\frac{\rho}{2}$
t	$\begin{pmatrix} \frac{1}{2}\rho & -\frac{1}{2} & \frac{1}{2}\rho \\ & -\rho & -\frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{36}+\frac{1}{8\pi}(3\sin^{-1}\frac{1}{2}\rho-\sin^{-1}\rho)+\frac{1}{8\pi^2}\{(\sin^{-1}\frac{1}{2}\rho)^2-(\sin^{-1}\rho)^2\}$
u	$\begin{pmatrix} \rho & \frac{1}{2}\rho \\ & \frac{1}{2} \end{pmatrix}$	$\frac{1}{6}+\frac{1}{4\pi}\sin^{-1}\frac{\rho}{2}+\frac{1}{4\pi}\sin^{-1}\rho.$
v	$\begin{pmatrix} \rho & -\frac{1}{2} & \frac{1}{2}\rho \\ & -\frac{1}{2}\rho & \frac{1}{2} \\ & & \frac{1}{2}\rho \end{pmatrix}$	$\frac{1}{18}+\frac{1}{8\pi}(\sin^{-1}\rho+\sin^{-1}\frac{1}{2}\rho)+\frac{1}{8\pi^2}\{(\sin^{-1}\rho)^2-(\sin^{-1}\frac{1}{2}\rho)^2\}$
w	$\begin{pmatrix} \rho & -\frac{1}{2} & -\frac{1}{2}\rho \\ & -\frac{1}{2}\rho & -\frac{1}{2} \\ & & \rho \end{pmatrix}$	$\frac{1}{36}+\frac{1}{4\pi}(\sin^{-1}\rho-\sin^{-1}\frac{1}{2}\rho)+\frac{1}{4\pi^2}\{(\sin^{-1}\rho)^2-(\sin^{-1}\frac{1}{2}\rho)^2\}$
x	(ρ)	$\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\rho$

APPENDIX 3

Integrals as series expansions

$$A \quad \int_0^x \sin^{-1}\left(\frac{\sin 2x}{\sqrt{(1+2\cos 2x)}}\right)dx = \sqrt{3}\left(\frac{x^2}{3}+\frac{x^4}{3^3}+\frac{2^2}{3^3.5}x^6+\frac{71}{3^4.5.7}x^8+\frac{2.4393}{3^7.5^2.7}x^{10}+\frac{2^3.11^2}{3^5.5^2.7}x^{12}+\ldots\right)$$

$$B \quad \int_0^x \sin^{-1}\left(\frac{\sin x}{\cos 2x}\right)dx = \frac{x^2}{2}+\frac{x^4}{2}+\frac{2}{3}x^6+\frac{11}{10}x^8+\frac{1999}{3^3.5.7}x^{10}+\frac{2^6.1009}{3^4.5^2.7}x^{12}+\ldots$$

$$C \quad \int_0^x \sin^{-1}\left(\frac{\sin x}{1+2\cos 2x}\right)dx = \frac{x^2}{2.3}+\frac{2^3}{3^4}x^4+\frac{2^4}{3^5}x^6+\frac{2^3.67}{3^7.5}x^8+\frac{2^4.15{,}437}{3^{11}.5.7}x^{10}+\frac{2^4.67{,}217}{3^{11}.5^2.7}x^{12}+\ldots$$

$$D \quad \int_0^x \sin^{-1}\left(\frac{\sin x}{3}\right)dx = \frac{x^2}{2.3}-\frac{x^4}{3^4}+\frac{x^8}{2.3^7.5}-\frac{2^2}{3^{11}.5.7}x^{10}-\frac{2}{3^{11}.5^2}x^{12}+\ldots$$

$$E \quad \int_0^x \sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x-1}{3(\cos 2x+1)}}\right)dx = \frac{x^2}{2.3}-\frac{2.5}{3^4}x^4-\frac{2^2.7}{3^5}x^6-\frac{2.5^2.7}{3^7}x^8 - \frac{2^2.437{,}671}{3^{11}.5.7}x^{10}-\frac{2^2.43{,}133}{3^{10}.5}x^{12}+\ldots$$

$$F \quad \int_0^x \sin^{-1}\left(\sin x\frac{2\cos 2x+3}{2\cos 2x+1}\right)dx = \frac{5}{2.3}x^2+\frac{2^2.7}{3^4}x^4+\frac{2^3.13}{3^5}x^6+\frac{2^2.1849}{3^7.5}x^8 + \frac{2^3.143{,}723}{3^{11}.5}x^{10}+\frac{2^3.11{,}014{,}679}{3^{11}.5^2.7}x^{12}+\ldots$$

$$G \quad \int_0^x \sin^{-1}\left(\frac{\sin x}{\sqrt{3}}\right)dx = \sqrt{3}\left(\frac{x^2}{2.3}-\frac{x^4}{2^2.3^3}-\frac{x^6}{2^2.3^4.5}+\frac{x^8}{2^4.3^4.5.7}+\frac{73}{2^4.3^7.5^2.7}x^{10}+\frac{x^{12}}{2^5.3^5.5^2.7}+\ldots\right)$$

$$H \quad \int_0^x \sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x-1}{3\cos 2x}}\right)dx = \sqrt{3}\left(\frac{x^2}{2.3}-\frac{5}{2.3^3}x^4-\frac{49}{3^4.5}x^6-\frac{221}{2.3^4.7}x^8 - \frac{138{,}767}{3^7.5^2.7}x^{10}-\frac{917}{3^5.5}x^{12}+\ldots\right)$$

$$I \quad \int_0^x \sin^{-1}\left(\sin x\sqrt{\frac{2\cos 2x+1}{\cos 2x}}\right)dx = \sqrt{3}\left(\frac{x^2}{2}+\frac{x^4}{2.3}+\frac{31}{3^3.5}x^6+\frac{743}{2.3^3.5.7}x^8 + \frac{11{,}017}{3^4.5^2.7}x^{10}+\frac{24{,}217}{3^4.5^2.7}x^{12}+\ldots\right)$$

$$J \quad \int_0^x \sin^{-1}\left(\frac{\sin 2x}{2\sqrt{(\cos 2x)}}\right)dx = \frac{x^2}{2}+\frac{x^4}{2^3}+\frac{7}{2^4.3}x^6+\frac{343}{2^7.3.5}x^8+\frac{58{,}111}{2^8.3^3.5.7}x^{10}+\frac{719{,}813}{2^{10}.3^4.5^2}x^{12}+\ldots$$

$$K \quad \int_0^x \sin^{-1}\left(\frac{3\sin x-\sin 3x}{4\cos 2x}\right)dx = \frac{x^4}{2^2}+\frac{x^6}{2^2}+\frac{293}{2^6.3.5}x^8+\frac{2495}{2^5.3^3.7}x^{10}+\frac{1{,}444{,}997}{2^9.3^3.5^2.7}x^{12}+\ldots.$$

Biometrika (1961), **48**, 1 *and* 2, *p.* 29
Printed in Great Britain

Tests for rank correlation coefficients. II

BY E. C. FIELLER AND E. S. PEARSON
Statistical Advisory Unit, War Office, and University College London

(This paper must, I fear, be introduced with a short explanatory statement. Before his untimely death on 1 December 1960, Edgar Fieller had organized and seen completed all the heavy calculations for the distributions of the coefficient r_F and its transform z_F. Publication was delayed because as far as he and I were aware there were no theoretical results for the mean and variance of r_F. However, since the empirical results establish clearly the utility of z_F from the practical point of view it seems now desirable to put these results on record, and to include in the same paper the new results for the variance of r_S. For the arrangement of the material and its discussion I must accept responsibility, but it should be stated that of the three authors of part I of this series it was Fieller who had from the beginning insisted on the inclusion of the r_F and z_F coefficients in the investigation. He was undoubtedly right in pressing this point. E.S.P.)

1. INTRODUCTION

In the first paper in this series, which will be described as part I (Fieller, Hartley & Pearson, 1957), 25,000 sets of correlated random normal deviates were used to examine the distributions of Spearman's and Kendall's rank correlation coefficients, r_S and r_K, in samples of $n = 10$, 30 and 50 observations. In particular, it was shown that, as in the case of the product moment correlation coefficient r_{xy}, R. A. Fisher's transformation

$$z = \tanh^{-1} r = \tfrac{1}{2} \log_e \frac{1+r}{1-r} \tag{1}$$

had a remarkable normalizing property when applied both to r_S and r_K. In this respect, and also as regards power of discrimination, it appeared that there was little to choose between these two coefficients. It was also pointed out that the conclusions reached would apply not only to rankings generated by sampling from a bivariate normal parent having correlation ρ, but to a much wider class of parental distributions which have the property of being convertible to the normal by monotonic transformations applied to the marginal distributions.

The present part II contains: (*a*) A short discussion of the more exact results for the variance of Spearman's r_S derived from the work of David & Mallows (1961). (*b*) Results for the Fisher–Yates coefficient r_F which had been incomplete in 1957. (*c*) Some comparison of the distributions of the three transformed coefficients z_S, z_K and z_F. (*d*) A numerical illustration.

2. THE VARIANCE OF SPEARMAN'S COEFFICIENT r_S

The results given in Table 1 have been calculated from equation (X) for var (r_S) given by David & Mallows (1961) on p. 22 of the preceding paper. The computation involved was carried out by Barbara Snow in the Department of Statistics, University College; as a check on arithmetic she also used the David & Mallows alternative formula (Z) involving the expansion of $S_1 = \sin^{-1}\rho$ and $S_2 = \sin^{-1}\frac{1}{2}\rho$ in powers of ρ. From the computing point of view it is of interest to note that the results based on (X) and (Z) agreed to the five decimal places tabled, except for the case $\rho = 0{\cdot}9$ where there were differences of 2 units ($n = 10$) and 1 unit ($n = 30, 50$) in the last place.

Table 1. *Theoretical values of* var (r_S)

ρ	$n = 10$	$n = 30$	$n = 50$
0·1	0·10974	0·03399	0·02010
0·2	·10564	·03250	·01920
0·3	·09886	·03006	·01771
0·4	·08947	·02672	·01568
0·5	·07762	·02258	·01317
0·6	0·06353	0·01777	0·01028
0·7	·04762	·01254	·00715
0·8	·03058	·00727	·00404
0·9	·01363	·00260	·00137

Since the α, β and γ terms of (X) consist of expansions in powers of S_1 and S_2, the curtailment of these series at the 12th power must involve some inaccuracy. No precise estimate of this error is possible without further investigation. David & Mallows considered the cases of $\rho = 0{\cdot}5$ and $0{\cdot}6$ numerically. If we take $\rho = 0{\cdot}9$, it is found that it is the β-term which converges most slowly. For $n = 10, \rho = 0{\cdot}9$ the full contributions to var (r_S) from this term are:

From terms in:	S_2^2	S_2^4	S_2^6	S_2^8	S_2^{10}	S_2^{12}	
	$+0{\cdot}062\,465$	9 552	2 635	1 126	388	183	(all positive)

Without exploring further, one might hazard a guess that in this case the calculated variance may be too small by 2 or 3 units in the 4th decimal place. For $n = 30$ and 50, the error should be no more than 1 or 2 units in the 5th place, even when $\rho = 0{\cdot}9$. With this caution, it is justifiable to table results to 5 decimal places throughout.

The figures for var (r_S) in Table 1 may be compared with those given in Table 1 of part I. It will then be noted that:

(*a*) As expected, the Kendall approximation breaks down for high ρ and low n.

(*b*) The results of our experimental sampling conform with the improved theoretical values.

(*c*) The smoothing of the experimental values, used in the further analysis, was on the whole successful in providing estimates of the true values.

With regard to (*c*), we had used these smoothed values to calculate approximations to the mean and variance of z_S and z_K, obtained from

$$\mathscr{E}(z) = \tanh^{-1}\bar{r} + \frac{\bar{r}\,\mathrm{var}\,(r)}{(1-\bar{r}^2)^2}, \tag{2}$$

$$\mathrm{var}\,(z) = \frac{\mathrm{var}\,(r)}{(1-\bar{r}^2)^2}. \tag{3}$$

The substitution of the true value of var (r_S) therefore makes no appreciable alteration in the figures given in Tables 2 and 3 of part I. It may therefore still be concluded that:

(i) The two terms of the expansion of equation (2) are adequate to give the expected value of z_S to all necessary accuracy, if $n \geqslant 10, \rho \leqslant 0{\cdot}9$.†

† The changes in var (r_s) from the earlier smoothed values bring the observed and theoretical values of $\bar{z}_s$ closer together.

(ii) The single term of equation (3) is not adequate to give var (z_S). For example we find:

		$n = 10$	$n = 30$	$n = 50$
Mean value of var (z_s) for $\rho = 0{\cdot}1 - 0{\cdot}8$†	From (3), using smoothed var (r_s)	0·1224	0·0370	0·0219
	From (3), using true var (r_s)	·1216	·0374	·0221
	From sampling experiment	·1487	·0388	·0228
	From empirical estimate, $1{\cdot}060/(n-3)$	·1514	·0392	·0226

There appears to be no reason, therefore, to alter the purely empirical suggestion of part I, namely to take

$$\text{var}\,(z_S) = \frac{1{\cdot}060}{n-3} \quad \text{or} \quad \sigma_z = \frac{1{\cdot}03}{\sqrt{(n-3)}}.$$

3. The distributions of the Fisher–Yates coefficient r_F and its transform z_F

(3·1) *Definition of* r_F

In the notation of part I, we have n pairs of associated rankings

$$u_1,\ u_2,\ \ldots,\ u_n \quad \text{and} \quad v_1,\ v_2,\ \ldots,\ v_n,$$

where the integers u_i $(i = 1, 2, \ldots, n)$ may be taken in ascending order $1, 2, \ldots, n$ and the v_i are a permutation of these integers. If we then attach to both these rankings the appropriate normal order statistics $\xi(i|n)$, i.e. the expected values of the ith largest standardized deviate in a sample of size n from a normal population, the Fisher–Yates coefficient (1938, pp. 14–15) is

$$r_F = \sum_{i=1}^{n} \xi(i|n)\,\xi(v_i|n) \Big/ \sum_{i=1}^{n} \xi^2(i|n). \tag{4}$$

Tables of $\xi(i|n)$ have been given in several publications (e.g. Fisher & Yates, 1938, Table XX; Pearson & Hartley, 1954, Table 28; Harter, 1961); Fisher & Yates (1938, Table XXI) give the expressions $\sum_i \xi^2(i|n)$ for $n = 2\,(1)\,50$.

(3·2) *Test of the hypothesis that* $\rho = 0$

Whatever be the form of the parent distribution, it follows from Pitman (1937) that if the two variables are independent

$$\mathscr{E}(r_F^2) = \frac{1}{n-1}, \tag{5}$$

$$\mathscr{E}(r_F^4) = \frac{1}{(n-1)^2}\left\{\frac{3(n-1)}{n+1} + \frac{(n-2)\,(n-3)}{n(n^2-1)}\left(\frac{k_4}{k_2^2}\right)^2\right\}, \tag{6}$$

where k_2 and k_4 are the 2nd and 4th cumulants of the $\xi(i|n)$, i.e.

$$\left.\begin{aligned} k_2 &= \frac{1}{n-1}\sum_i \xi^2(i|n), \\ k_4 &= \frac{n}{(n-1)\,(n-2)\,(n-3)}\left\{(n+1)\sum_i \xi^4(i|n) - \frac{3(n-1)}{n}\left[\sum_i \xi^2(i|n)\right]^2\right\}. \end{aligned}\right\} \tag{7}$$

† As before, in taking averages values for $\rho = 0{\cdot}9$ have been omitted.

The odd moments of r_F are of course zero. The second term within the braces in (6) will tend rapidly to zero as n increases. For example, when $n = 10$ it is found that

$$k_4/k_2^2 = -0{\cdot}4911$$

so that the two terms inside the braces become 2·4545 and 0·0136, respectively. Thus the second term is effectively negligible even when $n = 10$. As Pitman pointed out, $1/(n-1)$ and $3(n-1)/(n+1)$ are the variance and mean of the distribution

$$p(r) = \text{constant}\,(1-r^2)^{\frac{1}{2}(n-4)} \quad (-1 \leqslant r \leqslant 1).$$

It follows that without much loss of accuracy, in testing whether ρ differs from zero we may treat r_F as though it were the product moment correlation, r_{xy}, of n pairs of independently distributed normal variables. This result was implied, though not specifically stated, on pp. 14–15 of Fisher & Yates's Introduction (1938) and was again referred to by Hoeffding (1951, p. 88).

When $\rho \neq 0$, the distributions of r_{xy} and r_F will not be as similar; for example

$$\mathscr{E}(r_F) < \mathscr{E}(r_{xy}).$$

(3·3) *The distributions of r_F and their mean and variance*

As in the case of Spearman and Kendall coefficients, values of r_F defined by equation (4) were computed for each of the 2500 samples of 10, 833 samples of 30 and 500 samples of 50. The resulting distributions are not published here, but are available should any statistician wish to make use of them in exploratory work. Table 2 shows the mean and variance of each of the 27 distributions, $n = 10, 30, 50$; $\rho = 0{\cdot}1\,(0{\cdot}1)\,0{\cdot}9$.

Table 2. *Mean and variance of r_F. Experimental values*

ρ	$n = 10$		$n = 30$		$n = 50$	
	Mean	Variance	Mean	Variance	Mean	Variance
0·1	0·0868	0·10639	0·0965	0·03278	0·0959	0·01887
0·2	·1740	·10269	·1906	·03285	·1904	·02143
0·3	·2640	·09833	·2857	·03032	·2909	·01866
0·4	·3550	·09018	·3857	·02506	·3896	·01512
0·5	·4380	·07778	·4719	·02152	·4806	·01228
0·6	0·5387	0·06063	0·5725	0·01603	0·5837	0·009582
0·7	·6415	·04221	·6760	·009859	·6863	·005465
0·8	·7298	·02952	·7766	·005240	·7856	·002807
0·9	·8457	·01161	·8796	·001954	·8873	·000896

As far as is known, the true moments of r_F for any ρ have not been found although there is little doubt that with sufficient effort they could be obtained following the approach of David & Mallows (1961). We have therefore proceeded to an empirical examination of the distribution of z_F, each experimental value of r_F being converted by the transformation of equation (1) and the results tabled.

Table 3. *Comparison of observed and theoretical values for the mean and variance of* z_F

ρ	Smoothed		$\frac{\text{var}(r_F)}{(1-\bar{r}_F^2)^2}$	Experimental var(z_F)	Smoothed			Experimental $\bar{z}_F$	$\tanh^{-1}\rho$
	$\bar{r}_F$	var($\bar{r}_F$)			$\tanh^{-1}\bar{r}_F$	$\frac{\bar{r}_F \text{var}(r_F)}{(1-\bar{r}_F^2)^2}$	Cols. (6)+(7)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				$n = 10$					
0·1	0·0868	0·1076	0·109	0·1368	0·087	0·009	0·096	0·0966	0·1003
0·2	·1743	·1034	·110	·1370	·176	·019	·195	·1971	·2027
0·3	·2626	·0980	·113	·1426	·269	·030	·299	·3041	·3095
0·4	·3518	·0897	·117	·1482	·367	·041	·408	·4185	·4236
0·5	·4423	·0773	·119	·1465	·475	·053	·528	·5268	·5493
0·6	0·5346	0·0619	0·121	0·1410	0·596	0·065	0·661	0·6709	0·6931
0·7	·6300	·0450	·124	·1349*	·741	·078	·819	·8401*	·8673
0·8	·7320	·0282	·131	·1543*	·933	·096	1·029	1·0290*	1·0986
0·9	·8460	·0121	·150	·1654*	1·242	·127	1·369	1·3637*	1·4722
				$n = 30$					
0·1	0·0950	0·03370	0·0343	0·0360	0·095	0·003	0·099	0·1004	
0·2	·1902	·03260	·0351	·0379	·193	·007	·199	·2001	
0·3	·2854	·03025	·0359	·0384	·294	·010	·304	·3050	
0·4	·3810	·02600	·0356	·0359	·401	·014	·415	·4193	
0·5	·4770	·02125	·0356	·0382	·519	·017	·536	·5300	
0·6	0·5739	0·01585	0·0352	0·0358	0·653	0·020	0·674	0·6706	
0·7	·6724	·01000	·0333	·0331	·815	·022	·837	·8432	
0·8	·7740	·00525	·0327	·0342	1·030	·025	1·056	1·0625	
0·9	·8804	·00195	·0386	·0397	1·377	·034	1·411	1·4085	
				$n = 50$					
0·1	0·0967	0·02035	0·0207	0·0199	0·097	0·002	0·099	0·0988	
0·2	·1935	·01970	·0213	·0240	·196	·004	·200	·1969	
0·3	·2905	·01825	·0218	·0228	·299	·006	·305	·3063	
0·4	·3880	·01565	·0217	·0217	·409	·008	·418	·4198	
0·5	·4859	·01255	·0215	·0216	·531	·010	·541	·5338	
0·6	0·5845	0·00910	0·0210	0·0221	0·669	0·012	0·681	0·6809	
0·7	·6841	·00565	·0200	·0204	·837	·014	·851	·8541	
0·8	·7852	·00281	·0191	·0195	1·059	0·015	1·074	1·0745	
0·9	·8890	·00090	·0205	·0198	1·417	·018	1·435	1·4264	

* For $n = 10$, the few infinite values of z have been ignored for $\rho = 0{\cdot}7$, $0{\cdot}8$ and $0{\cdot}9$.

To improve the approximations, (2) and (3), to $\mathscr{E}(z_F)$ and var (z_F), the observed values of r_F and var (r_F) were smoothed. This was done graphically, making use of the known values of these statistics corresponding to the terminal values $\rho = 0$ and 1. The results are given in cols. (2) and (3) of Table 3.

(3·4) *The variance of* z_F

Clearly the most important function of the z-transformation is to provide approximate normality and a variance as nearly as possible independent of the population ρ. We shall consider first the variance. It is clear from a comparison of cols. (4) and (5) of Table 3 that, as in the cases of z_S and z_K considered in part I, the approximation of equation (3) is far from adequate when $n = 10$ and still giving too small values when $n = 30$. As the true mean and variance of r_F are at present undetermined, this is of no great consequence.

What is of much more importance is that the experimental values of var (z_F) have an average of nearly $1/(n-3)$, i.e. the same value as is used for the transformed product moment correlation coefficient, z_{xy}. Thus we find:

		$n = 10$	$n = 30$	$n = 50$
Average of estimates of var (z_F) for $\rho = 0{\cdot}1 - 0{\cdot}8$	Approx. (3) smoothed	0·118	0·0347	0·0209
	Experiment	·143	·0362	·0215
	$1/(n-3)$	·143	·0370	·0213

Certainly the variance will not be exactly constant, but the sampling error obscures such trend with ρ as is present. We may note that the variance of the product moment z_{xy} as given by Gayen's (1951) amendment to the original expression of Fisher takes the following values:

ρ \ n	10	30	50
0·1	0·141	0·0370	0·0213
0·5	·139	·0368	·0212
0·9	·134	·0365	·0211

(3·5) *The expectation of* z_F

Column (9) of Table 3 gives the experimental values of $\bar{z}_F$; the standard errors are, approximately: for $n = 10$, 0·008; for $n = 30$, 50, 0·007. Cols. (6), (7) and (8) were obtained by inserting the smoothed values of $\bar{r}_F$ and var (r_F) from cols. (2) and (3) into the approximation to $\mathscr{E}(z_F)$ of equation (2). It seems probable that this expression would be adequate were the true values of $\bar{r}_F$ and var (r_F) known.

Because of the very small change in $\mathscr{E}(z_F|\rho, n)$ with n, a knowledge of this function is not in general required in using tests of significance. As, however, it might be needed in connexion with estimation, it was thought that a purely empirical formula (with no theoretical significance) would be worth recording. For simplicity this will need to be expressed directly in terms of ρ and n rather than the unknown $\bar{r}_F$ and var (r_F).

Column (10) of the top section of Table 3 gives values of $\tanh^{-1}\rho$. It is seen that the experimental means lie between these values and $\tanh^{-1}\bar{r}_F$. This suggested that an expression of the form $\tanh^{-1}\rho'$, with $\rho' < \rho$, might be used. After several attempts, it was found that a good approximation could be obtained by putting $\rho' = \rho\{1 - 0{\cdot}6/(n+8)\}$, i.e. taking

$$\mathscr{E}(z_F) \sim \tanh^{-1}\rho\left(1 - \frac{0{\cdot}6}{n+8}\right). \tag{8}$$

Table 4 compares this approximation with the figures obtained previously by inserting smoothed values of $\bar{r}_F$ and $\text{var}(r_F)$ into equation (2). The agreement certainly appears adequate for practical use and is likely to hold to the same order of approximation for *intermediate values* of ρ and n.

Table 4. *Comparison of* (i) $\tanh^{-1}(\rho') = \tanh^{-1}\left\{\rho\left(1-\frac{0{\cdot}6}{n+8}\right)\right\}$ *and* (ii) *the approximation* $\mathscr{E}(z_F) = \tanh^{-1}\bar{r}_F + \bar{r}_F\,\text{var}(r_F)/(1-\bar{r}_F^2)^2$ *(smoothed)*

ρ	$n = 10$		$n = 30$		$n = 50$	
	$\tanh^{-1}(\rho')$	$E(z_F)$	$\tanh^{-1}(\rho')$	$E(z_F)$	$\tanh^{-1}(\rho')$	$E(z_F)$
0·1	0·097	0·096	0·099	0·099	0·099	0·099
0·2	·196	·195	·199	·199	·201	·200
0·3	·299	·299	·304	·304	·306	·305
0·4	·408	·408	·416	·415	·419	·418
0·5	·527	·528	·539	·536	·542	·541
0·6	0·662	0·661	0·673	0·674	0·684	0·681
0·7	·823	·819	·846	·837	·853	·851
0·8	1·029	1·029	1·065	1·056	1·076	1·074
0·9	1·333	1·369	1·402	1·411	1·425	1·435

(3·6) *Shape of the distribution of* z_F

Values of the moment ratios $\sqrt{b_1}$ and b_2 were calculated for all distributions except the three containing infinite values of z (i.e. $n = 10$; $\rho = 0{\cdot}7, 0{\cdot}8, 0{\cdot}9$). They are shown in Tables 5 and 6, together with the corresponding values for z_S.† The large sample approximations to the standard errors of $\sqrt{b_1}$ and b_2 in sampling from a *normal distribution*, namely $\sqrt{(6/N)}$ and $\sqrt{(24/N)}$, give

	$N = 2500$ for $n = 10$	$N = 833$ for $n = 30$	$N = 500$ for $n = 50$
$\sigma(\sqrt{b_1})$	0·049	0·085	0·110
$\sigma(b_2)$	0·098	0·170	0·219

Except perhaps in the case of samples of $n = 10$, there is no consistent evidence of asymmetry in the distributions of either z_S or z_F. On the other hand the distributions are clearly leptokurtic, with $\beta_2 > 3$. This effect seems to be appreciably the same for both z_S and z_F and is definitely rather greater than that for the transformed product moment coefficient z_{xy}. The β_2 values for the last, averaged for $\rho = 0{\cdot}1$–$0{\cdot}8$, are shown at the bottom of Table 6.

To obtain some idea of the effect of this amount of kurtosis on tests of significance (based on the assumption of normality), we may suppose that the distribution of z_F is represented by a symmetrical Pearson curve (a type VII or t-distribution) with the same β_2 values.

† (It appears that the values of these statistics were not completely calculated by E.C.F. for Kendall's coefficient. E.S.P.)

Table 5. *Values of $\sqrt{b_1}$ for distributions of z*

ρ	z_S (experiment)			z_F (experiment)			z_{av} for $n = 10$ (theory)*
	$n = 10$	$n = 30$	$n = 50$	$n = 10$	$n = 30$	$n = 50$	
0·1	−0·17	0·03	0·01	−0·11	0·00	−0·04	0·000
0·2	·02	− ·07	− ·13	− ·02	− ·04	− ·12	·000
0·3	·09	− ·03	·05	·06	− ·04	− ·03	·001
0·4	·18	− ·02	·10	·14	− ·17	·01	·002
0·5	·21	·26	·24	·12	·14	·12	·005
0·6	0·17	−0·08	−0·07	0·11	−0·16	−0·09	0·008
0·7	—	− ·07	·12	—	− ·07	·08	·013
0·8	—	·07	·04	—	·09	·05	·019
0·9	—	·17	·00	—	·22	·04	·027

* From the Fisher–Gayen formula.

Table 6. *Values of b_2 for distributions of z*

ρ	z_S (experiment)			z_F (experiment)		
	$n = 10$	$n = 30$	$n = 50$	$n = 10$	$n = 30$	$n = 50$
0·1	3·641	3·398	3·024	3·427	3·360	3·242
0·2	3·487	3·103	2·940	3·325	3·218	2·814
0·3	3·308	3·302	3·335	3·237	3·377	2·946
0·4	3·353	3·151	2·996	3·323	3·092	3·017
0·5	3·543	3·297	3·163	3·692	3·228	3·010
0·6	3·267	3·163	3·297	3·397	3·316	3·326
0·7	—	3·224	3·043	—	3·265	2·791
0·8	—	3·036	2·641	—	2·925	2·791
0·9	—	3·083	2·437	—	3·125	2·644
Mean $\rho = 0·1 - 0·8$*	3·433	3·209	3·055	3·400	3·223	2·992
Mean for z_{av}†	3·274	3·074	3·043	—	—	—

* For $n = 10$, the mean is only for $\rho = 0·1 - 0·6$.
† From the Fisher–Gayen formula.

Table 42 of Pearson & Hartley (1954) shows the following modification of standardized percentage points, according to β_2 (when $\beta_1 = 0$):

		5%	2·5%	1·0%	0·5%	0·25%†
Normal curve,	$\beta_2 = 3{\cdot}0$	1·64	1·96	2·33	2·58	2·81
Type VII	$\beta_2 = 3{\cdot}2$	1·64	1·97	2·37	2·65	2·91
	$\beta_2 = 3{\cdot}4$	1·64	1·98	2·40	2·71	3·00

The effect only becomes of importance at extreme significance levels and, presumably, has never been taken into account in the practical application of the z_{xy} transformation in small samples.

Normal curves were fitted to the distributions and the values of χ^2 obtained were of about the same average size as those given for z_S and z_K in Table 7 of part I.

4. The sensitivity of rank correlation measures to changes in ρ

In § 5 of part I brief consideration was given to the question of power. It was pointed out that the power of discrimination of any one of these alternative correlation measures would depend on the rapidity with which its sampling distributions 'drew clear' of one another as the population value of ρ changed. It was suggested that a rough measure of local sensitivity would be given by the ratios $(\bar{z}_2 - \bar{z}_1)/\sqrt{(s^2_{z_1} + s^2_{z_2})}$ of

(*a*) the difference between pairs of consecutive entries for $\bar{z}_F$ in col. (9) of Table 3 (or corresponding values for $\bar{z}_S$ and $\bar{z}_K$ in part I), and

(*b*) the square root of the sum of the corresponding pair of variances from col. (5) of the same table.

Table 7. *Sensitivity ratios* $(\bar{z}_1 - \bar{z}_2)/\sqrt{(s^2_{z_1} + s^2_{z_2})}$ *for* z_F

ρ_1, ρ_2	$n = 10$	$n = 30$	$n = 50$
0·1, 0·2	0·192	0·367	0·471
0·2, 0·3	·202	·380	·506
0·3, 0·4	·212	·419	·538
0·4, 0·5	·199	·407	·548
0·5, 0·6	0·269	0·517	0·704
0·6, 0·7	·322	·658	·840
0·7, 0·8	·351	·845	1·10
0·8, 0·9	·591	1·35	1·78

Table 8 of part I compared these sensitivity ratios for the z transforms of the product moment and of Spearman's and Kendall's rank correlation coefficients. Table 7 gives similar ratios for the Fisher–Yates coefficient. On comparing the two tables it will be found that with very few exceptions, which are probably the result of sampling fluctuations,‡ the sensitivity ratio for z_F lies between that for z_{xy} and those for z_S and z_K. In so far as the variances of the z's are nearly constant, the sensitivity in detecting differences in ρ greater than 0·1 will be found by adding two or more of the consecutive local ratios. The superiority

† Results for 0·25% from a table in preparation by N. L. Johnson & E. Nixon.

‡ As in part I, the ratios are calculated from the unsmoothed experimental means and variances.

of z_F over z_S and z_K will then stand out further still. It follows from this evidence that for the population model considered a test based on the Fisher–Yates coefficient is more likely to detect differences in the population ρ, if they exist, than tests based on the other two rank coefficients. This empirical result confirms what would be expected on theoretical grounds.

5. A NUMERICAL EXAMPLE

In Table 8 there are shown for each of four samples, n_t pairs of associated rankings ($t = 1, 2, 3, 4$), where $n_1 = 15$, $n_2 = 12$, $n_3 = 19$ and $n_4 = 14$. Assuming that the underlying ranked variables x_{ti}, y_{ti} follow four bivariate normal distributions, are the results consistent with the hypothesis that these parent distributions have a common coefficient of correlation ρ? If the parent distributions are not bivariate normal, it is assumed that they each

Table 8. *Comparison of four tests for heterogeneity of correlation*

$i = u_t$	Values of v_t					
	1st sample $n_1 = 15$	2nd sample $n_2 = 12$	3rd sample $n_3 = 19$	4th sample $n_4 = 14$		
1	9	2	3	3		
2	11	3	8	1		
3	4	6	1	4		
4	7	4	9	2		
5	2	1	2	7		
6	5	12	10	6		
7	1	9	5	9		
8	10	8	6	8		
9	13	5	18	5		
10	14	10	4	13		
11	6	11	19	10		
12	15	7	11	12		
13	12	—	16	11		
14	3	—	17	14		
15	8	—	15	—		
16	—	—	13	—		
17	—	—	7	—		
18	—	—	14	—		
19	—	—	12	—		
r_S	0·211	0·622	0·583	0·894		
r_K	·143	·424	·368	·714		
r_F	·164	·552	·518	·875		
r_{xy}	·161	·500	·562	·864	$\bar{z}$	χ^2
z_S	0·214	0·729	0·666	1·444	0·743	8·36
z_K	·144	·453	·387	0·896	·454	7·04
z_F	·166	·621	·573	1·354	·659	8·36
z_{xy}	·163	·550	·636	1·308	·656	7·70

have the property of being convertible to the normal by monotonic transformations applied to their margins. We should then be testing whether the distributions, after transformation, had a common ρ.

We can now proceed by using any one of the three rank correlation coefficients r_S, r_K or r_F, applying the z-transformation to the coefficients calculated for each sample and then determining

$$\chi^2 = \sum_{i=1}^{4} w_i(z_i - \bar{z})^2 \quad \text{with } \nu = 3 \text{ degrees of freedom,} \tag{9}$$

where

$$\bar{z} = \sum_{i=1}^{4} w_i z_i \Big/ \sum_{i=1}^{4} w_i \tag{10}$$

and w_i is the reciprocal of the variance appropriate to the type of coefficient chosen. That is to say, using the empirical results:

for Spearman's coefficient	$w_i = (n_i - 3)/1{\cdot}060$,
for Kendall's coefficient	$w_i = (n_i - 4)/0{\cdot}437$,
for the Fisher–Yates' coefficient	$w_i = n_i - 3$.

At the bottom of Table 8, the correlation coefficients are shown with their z-transforms and also the mean z of equation (10) and the χ^2 values of equation (9). The rankings were in fact based on samples of 15, 12, 19 and 14 taken from Fieller, Lewis & Pearson's *Tables of Correlated Normal Deviates* (1955), the parent correlations being different, namely

$$\rho_1 = 0{\cdot}1, \quad \rho_2 = 0{\cdot}3, \quad \rho_3 = 0{\cdot}6, \quad \rho_4 = 0{\cdot}7.$$

We have also, therefore, shown the four values of r_{xy}, and their z-transforms and the resulting weighted mean z and χ^2. For $\nu = 3$ degrees of freedom, the 10 % point of χ^2 is at 6·25, the 5 % at 7·81 and the 2·5 % at 9·35. All four values are therefore on the borderline of significance. The χ^2's for the Spearman and Fisher–Yates coefficients have come out with the top value of 8·36, followed by that for the product moment coefficient with 7·70 and Kendall's coefficient with 7·04.

It must, of course, be emphasized that the ordering of χ^2 in a single set of samples will not necessarily agree with the order of efficiency with which the tests would be expected to detect differences in the ρ-values in the long run. For given differences in ρ, we should expect that χ^2 based on z_{xy} would on the average be greatest, the χ^2 based on z_F coming second.

From the point of view of computation, the coefficients r_S and r_K are obtained most easily while r_{xy} involves the longest calculation; after this stage, the procedure is in all cases the same apart from the difference in weights.

6. Discussion

The results discussed in the two papers of this series apply to rankings generated by sampling either from (*a*) a bivariate normal distribution, or (*b*) a class of bivariate distributions which are convertible to the normal form by monotonic transformations applied to the margins. In case (*b*), ρ is the coefficient of correlation in the transformed, not the original distribution. The properties of the class under (*b*) which contain a wide range of skew bivariate forms have not been extensively explored (but see Johnson, 1949).

On theoretical grounds it would be expected that when dealing with a bivariate normal parent, rank order tests based on r_F would be more powerful than those based on r_S. The coefficient r_K based on paired comparisons falls in a rather different category.

By making use of the z-transformation, all three coefficients are put into a simple comparable form. In all three cases the distributions of z have variances nearly independent of ρ, are closely symmetrical but slightly leptokurtic.

The advantage of the Fisher–Yates coefficient z_F over the other two lies in its somewhat greater power of discrimination, and in the fact that the sampling variance of z_F seems to have nearly the same value as that of z_{xy}, i.e. $1/(n-3)$—an easy expression to remember. A disadvantage, which certainly carries *some* practical weight is that a table or tables are needed to calculate r_F.† Thus the test does not fall within the class of quick methods which, to quote a description once used by Student, can be worked out in a railway train on the back of an envelope.

As in connection with part I, the authors have been greatly indebted to Miss M. U. Thomas, Mrs Esmé Hill and Mr T. Vickers for extensive help in computation.

REFERENCES

DAVID, F. N. & MALLOWS, C. L. (1961). The variance of Spearman's 'rho' in normal samples. *Biometrika*, **48**, 19–28.

FIELLER, E. C., HARTLEY, H. O. & PEARSON, E. S. (1957). Tests for rank correlation coefficients. I. *Biometrika*, **44**, 470–81.

FIELLER, E. C., LEWIS, T. & PEARSON, E. S. (1955). Correlated random normal deviates. *Tracts for Computers*, no. XXXVI. Cambridge University Press.

FISHER, R. A. & YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research* (5th edition, 1957). Edinburgh: Oliver and Boyd.

GAYEN, A. J. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, **51**, 219–47.

HARTER, H. L. (1961). Expected values of normal order statistics. *Biometrika*, **48**, 151–165.

HOEFFDING, W. (1951). Optimum non-parametric tests. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 83–92. University of California Press.

JOHNSON, N. L. (1949). Bivariate distributions based on simple transformation systems. *Biometrika*, **36**, 297–304.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*. Cambridge University Press.

PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient. *J. R. Statist. Soc.* Suppl. **4**, 225–32.

† A table is of course required in all cases to convert r to z.

Biometrika (1961), **48**, 1 *and* 2, *p.* 41
Printed in Great Britain

Some methods of constructing exact tests†

BY J. DURBIN
Research Technique Division, London School of Economics

1. INTRODUCTION

For testing the hypothesis that a set of observations $x_1, \ldots, x_n$ is a random sample from a distribution having a specified continuous distribution function $F(x)$ a variety of procedures are available, the most celebrated being Pearson's χ^2 test. The χ^2 test is flexible and easy to use but it possesses an element of arbitrariness in the choice of group boundaries; moreover, it does not have the desirable property of being an exact test in the sense of giving exact probabilities of rejection when the null hypothesis is true. The Kolmogorov test, which has been described and compared with related tests in an excellent article by Darling (1957), is free from these objections and is known to have good asymptotic power properties against alternatives specified in terms of distance between distribution functions. Nevertheless, in my experience its performance in practice with samples of moderate size has been disappointing in that it has frequently failed to give a significant result when the χ^2 test has done so. Some empirical results which illustrate this tendency will be given later in the paper.

The reason for the disappointing performance of the Kolmogorov test is not far to seek. If the alternative distribution is such that the difference between the hypothetical distribution function and the alternative distribution function is relatively large for some value of the variate, one would anticipate that Kolmogorov's test would have high power. Such would be the case, for example, if the two distributions differed only in location. If, however, the difference between the two distribution functions is nowhere large, as is quite possible if their means and variances are the same even though the two frequency functions differ markedly in shape, one would not anticipate that Kolmogorov's statistic would be a powerful discriminator between the two hypotheses.

Putting $u_j = F(x_j)\,(j = 1, \ldots, n)$, the hypothesis that $x_1, \ldots, x_n$ have distribution function $F(x)$ is equivalent to the hypothesis that $u_1, \ldots, u_n$ are $U(0, 1)$ variables, i.e. are uniformly distributed on the interval $(0, 1)$. Let $u_1, \ldots, u_n$ be ordered to give the ordered values $u_{(1)} \leqslant u_{(2)} \leqslant \ldots \leqslant u_{(n)}$. The first part of this paper shows how to transform $u_{(1)}, \ldots, u_{(n)}$ to a new set of values $w_1, \ldots, w_n$ which have the same joint distribution as $u_{(1)}, \ldots, u_{(n)}$ when the null hypothesis is true but which give a set of points tending to be more heavily concentrated towards the left-hand end of the $(0, 1)$ interval than towards the right-hand end under a wide class of alternatives. A variety of procedures can now be used for testing the hypothesis that $w_1, \ldots, w_n$ are ordered $U(0, 1)$ variables, special interest being attached in this paper to the one-sided Kolmogorov test. An indication is given in § 6 of applications to some problems other than those of goodness-of-fit, namely, to tests based on the periodogram in time-series analysis and to tests of the distribution of the intervals between successive events.

† This paper is based mainly on research done at the University of North Carolina, Chapel Hill, with the support of the Office of Naval Research. The principal results were presented in an invited address delivered to the Institute of Mathematical Statistics and the Biometric Society at New York, 26 April 1960. The work was completed and the paper written at Stanford University with the support of the National Science Foundation.

Because of wide applicability to problems in statistics and probability as well as for purely mathematical interest, distributions related to points scattered at random in an interval have been extensively studied in many different guises. No attempt will be made here to review the entire field; the following list of references is intended merely to indicate work which seems to me to be fairly closely connected with the subject matter of this paper. Among contributions concerned specifically with the random division of an interval we may mention those by Whitworth (1887), Moran (1947, 1951), Sherman (1950), Mauldon (1951), Darling (1953), Irwin (1955) and Barton & David (1955, 1956*a*, *b*). The following papers are concerned with goodness-of-fit aspects: K. Pearson (1933), Neyman (1937), E. S. Pearson (1939) and Darling (1957). Authors dealing with similar problems arising in the study of exponential distributions are Sukhatme (1936, 1937), Epstein & Tsao (1953), Epstein & Sobel (1954), Bartholemew (1957) and Epstein (1960). For applications to Poisson processes we mention the papers by Greenwood (1946), Maguire, Pearson & Wynn (1952, 1953), Barnard (1953), Cox (1955), and Bartholemew (1956). Relevant papers concerned primarily with the order statistics are those by Wilks (1948), Malmquist (1951), Renyi (1953) and van Dantzig (1954).

A severe limitation to the use hitherto of the Kolmogorov test has been that it has only been available for tests of simple hypotheses. Thus it has not been usable even for the classical problem of testing for normality, except as an approximation, since the hypothesis is a composite one involving two nuisance parameters, the population mean and variance. In § 6 a general method is proposed for eliminating nuisance parameters thereby enbling us to convert a composite hypothesis into a simple hypothesis. The idea underlying the method is to replace sufficient estimators for the unknown nuisance parameters by values picked at random from appropriate distributions. The price we pay for the elimination of the nuisance parameters is the introduction of an extraneous element of randomization. What we get in return is an exact knowledge of the probability of rejection.

2. RE-ORDERING OF INTERVALS

Suppose we have a sample of independently and identically distributed observations $x_1, \ldots, x_n$ and wish to test the hypothesis that they come from a distribution with continuous distribution function $F(x)$. When the null hypothesis is true the values $u_j = F(x_j)$ $(j = 1, \ldots, n)$ are independent $U(0, 1)$ variables. This implies that the n points determined by $u_1, \ldots, u_n$ are randomly scattered on the $(0, 1)$ interval. One would expect a departure from the null hypothesis to be indicated by a tendency for some of the intervals between adjacent points to be shorter and for some of the intervals between adjacent points to be longer than is found on the hypothesis of random scatter. This suggests that we take as our starting point a study of the relative lengths of the intervals between adjacent points.

Denoting the ordered u's by $u_{(1)}, \ldots, u_{(n)}$, these have the distribution

$$dP = n! du_{(1)} \ldots du_{(n)} \quad (0 \leqslant u_{(1)} \leqslant u_{(2)} \leqslant \ldots \leqslant u_{(n)} \leqslant 1). \tag{1}$$

Let $c_1 = u_{(1)}$, $c_j = u_{(j)} - u_{(j-1)}$ $(j = 2, \ldots, n)$ and $c_{n+1} = 1 - u_{(n)}$. Since the Jacobian is unity, the c's have the distribution

$$dP = n! dc_1 \ldots dc_n \quad \left(c_1 \geqslant 0, \ldots, c_{n+1} \geqslant 0, \sum_{j=1}^{n+1} c_j = 1\right). \tag{2}$$

A thorough discussion of this distribution has been given by Wilks (1948). Note that we avoid difficulties arising from the fact that the distribution of $c_1, \ldots, c_{n+1}$ is singular by

writing down the probability element for $c_1, \ldots, c_n$; the same device is used below for obtaining the distribution of the quantities $c_{(1)}, \ldots, c_{(n+1)}$ and $g_1, \ldots, g_{n+1}$.

Since we are interested in the relative magnitudes of the c's it is natural to consider the ordered c's, $c_{(1)} \leqslant c_{(2)} \leqslant \ldots \leqslant c_{(n+1)}$. Since there are $(n+1)!$ ways of permuting n objects from $n+1$ the joint distribution of $c_{(1)}, \ldots, c_{(n)}$ is, using (2),

$$dP = (n+1)!\, n!\, dc_{(1)} \ldots dc_{(n)} \quad \left(0 \leqslant c_{(1)} \leqslant \ldots \leqslant c_{(n+1)}, \sum_{j=1}^{n+1} c_{(j)} = 1\right). \tag{3}$$

We now transform $c_{(1)}, \ldots, c_{(n+1)}$ into a more manageable form by means of the transformation

$$g_j = (n+2-j)\,(c_{(j)} - c_{(j-1)}) \quad (c_{(0)} = 0;\ j = 1, \ldots, n+1). \tag{4}$$

The Jacobian is $1/(n+1)!$. Moreover, each $g_j \geqslant 0$ and

$$\sum_{j=1}^{n+1} g_j = \sum_{j=1}^{n+1} c_{(j)} = 1.$$

Consequently the distribution of $g_1, \ldots, g_n$ is

$$dP = n!\, dg_1 \ldots dg_n \quad \left(g_1 \geqslant 0, \ldots, g_{n+1} \geqslant 0, \sum_{j=1}^{n+1} g_j = 1\right). \tag{5}$$

Comparing (5) with (2) we see that the two distributions are identical. We therefore have the remarkable result that $g_1, \ldots, g_{n+1}$, which depend on the *ordered* intervals, have the same distribution as the *unordered* intervals $c_1, \ldots, c_{n+1}$. This result provides the key to the methods of test construction proposed in this and the following section. It depends essentially on the transformation (4) introduced by Sukhatme (1937) in the study of ordered exponentially distributed variables and used by Mauldon (1951) and Dwass (1959) in the study of random intervals.

Putting

$$w_r = \sum_{j=1}^{r} g_j, \tag{6}$$

it follows that $w_1, \ldots, w_n$ have the same distribution as the ordered $U(0,1)$ variables $u_{(1)}, \ldots, u_{(n)}$. Consequently any test procedure depending on $u_{(1)}, \ldots, u_{(n)}$, such as the Kolmogorov test, has the same properties under the null hypothesis when it is based on $w_1, \ldots, w_n$ as when it is based on $u_{(1)}, \ldots, u_{(n)}$.

It might be asked what has been gained by transforming to a set of values which have the same distribution as have the values we started with. The answer is that we hope to gain power. Some considerations which indicate why one would expect power to be increased in many situations will be discussed in the next two paragraphs.

The type of alternative we have in mind in this paper is one in which small intervals tend to be smaller and large intervals larger than on the null hypothesis. A way of making this idea precise is the following. Let primes denote values obtained on the alternative hypothesis, i.e. c'_j, $c'_{(j)}$, g'_j and w'_j are the values on the alternative hypothesis corresponding to values c_j, $c_{(j)}$, g_j and w_j on the null hypothesis. Suppose that whenever $c_i < c_j$ then $c'_i/c'_j < c_i/c_j$, i.e. ratios of larger to smaller intervals tend to become magnified on the alternative hypothesis. We shall show that this implies that $w'_j < w_j\ (j = 1, \ldots, n)$. For $c_{(j)} < c_{(j+1)}$, so that

$$c'_{(j)}/c'_{(j+1)} < c_{(j)}/c_{(j+1)} \quad (j = 1, \ldots, n)$$

(we ignore the event $c_{(j)} = c_{(j+1)}$ since this has probability zero). Thus

$$\frac{c'_{(j+1)} - c'_{(j)}}{c'_{(j)} - c'_{(j-1)}} = \frac{c'_{(j+1)}/c'_{(j)} - 1}{1 - c'_{(j-1)}/c'_{(j)}} > \frac{c_{(j+1)}/c_{(j)} - 1}{1 - c_{(j-1)}/c_{(j)}} = \frac{c_{(j+1)} - c_{(j)}}{c_{(j)} - c_{(j-1)}}.$$

Consequently

$$\frac{c'_{(j+1)} - c'_{(j)}}{c_{(j+1)} - c_{(j)}} > \frac{c'_{(j)} - c'_{(j-1)}}{c_{(j)} - c_{(j-1)}},$$

so that

$$\frac{g'_{j+1}}{g_{j+1}} > \frac{g'_j}{g_j} \quad (j = 1, \ldots, n),$$

i.e. g'_j/g_j is an increasing function of j. But

$$\sum_1^{n+1} g'_j = \sum_1^{n+1} g_j = 1.$$

Consequently $g'_1/g_1 < 1$ and $g'_{n+1}/g_{n+1} > 1$. These results together imply that there is a value r such that $g'_i \leqslant g_i$ for $i = 2, \ldots, r$ and $g'_i > g_i$ for $i > r$ where $1 \leqslant r < n+1$. Since $g'_1 < g_1$, for $j \leqslant r$ we have

$$w'_j = \sum_{i=1}^{j} g'_i < \sum_{i=1}^{j} g_i = w_j,$$

while for $j > r$ we have

$$w'_j = 1 - \sum_{i=j+1}^{n+1} g'_i < 1 - \sum_{i=j+1}^{n+1} g_i = w_j.$$

Thus $w'_j < w_j$ for all j.

It is remarkable that assuming only that ratios of larger to smaller intervals become greater as we proceed from the null to the alternative hypothesis we have been able to show that all values of the w's are diminished. In this situation the sample distribution function calculated from the w's is never less on the alternative hypothesis than on the null hypothesis. It is unlikely that the sample distribution function of the original u's would give as clear an indication of departure from the null hypothesis as that of the w's except perhaps for special types of departure such as those arising from change of location, for the detection of which a goodness-of-fit test would not normally be employed.

We observe in passing some important results which are, however, only of incidental interest for the purpose of this paper. First, we note that the quantities

$$\{c_j - c_{(1)}\}/\{1 - (n+1)\, c_{(1)}\} \quad (j = 1, \ldots, n+1),$$

deleting the one which is identically zero, are distributed like the intervals formed by $n-1$ random points on the $(0, 1)$ interval. This follows easily by considering the conditional distribution of $c_{(2)}, \ldots, c_{(n+1)}$ given $c_{(1)}$ and transforming back to the unordered c_j's. As a consequence we infer the general result that the quantities

$$\{c_j - c_{(r)}\}/\{1 - c_{(1)} - \ldots - c_{(r-1)} - (n+2-r)\, c_{(r)}\} \quad (j = 1, \ldots, n+1),$$

deleting the r values which are negative or identically zero, are distributed like the intervals determined by $n-r$ random points on the $(0, 1)$ interval. The quantities

$$\{c_{(j)} - c_{(j-1)}\}/\{1 - c_{(1)} - \ldots - c_{(j-2)} - (n+3-j)\, c_{(j-1)}\} \quad (c_{(0)} = 0;\ j = 1, \ldots, n)$$

are therefore independently distributed like the shortest of the intervals into which the interval $(0, 1)$ is divided by samples of $n, n-1, \ldots, 2$ random points, respectively. Now $(n+1)\, c_{(1)}$ is distributed like c_1 and c_1 has the distribution $dP = n(1-c_1)^{n-1}\, dc_1$ (Wilks, 1948),

whence $\{1-(n+1)c_{(1)}\}^n$ is a $U(0,1)$ variable. From this we deduce the general result that the quantities

$$\left\{1-\frac{(n+2-j)(c_{(j)}-c_{(j-1)})}{1-c_{(1)}-\ldots-c_{(j-2)}-(n+3-j)c_{(j-1)}}\right\}^{n+1-j} \qquad (c_{(0)}=0;\, j=1,\ldots,n)$$

are independent $U(0,1)$ variables.

3. Some new tests of goodness-of-fit

From (4) and (6) we have

$$w_j = c_{(1)}+\ldots+c_{(j-1)}+(n+2-j)c_{(j)} \quad (j=1,\ldots,n), \tag{7}$$

where $c_{(1)} \leqslant \ldots \leqslant c_{(n+1)}$ is the ordered set of intervals. The goodness-of-fit tests proposed in this section are all tests of the hypothesis that $w_1,\ldots,w_n$ are distributed as ordered $U(0,1)$ variables. Large numbers of such tests could be devised; here we mention only three which in addition to being exact have special features of interest. The first is included on account of its simplicity and because it requires only the widely available tables of the F-distribution. We call it the *modified median test* since it is based on the median of the w's. Consider the distribution of w_r. The probability element required is the probability that in a sample of n uniform variables $r-1$ are less than w_r, one is in the range w_r, w_r+dw_r and $n-r$ are greater than w_r, i.e.

$$dP = \frac{n!}{(r-1)!\,(n-r)!}\, w_r^{r-1}(1-w_r)^{n-r}\,dw_r \quad (0 \leqslant w_r \leqslant 1). \tag{8}$$

Thus

$$M_r = \frac{r}{n+1-r}\,\frac{1-w_r}{w_r}$$

has Fisher's variance-ratio distribution with $2(n+1-r)$, $2r$ degrees of freedom. The value of w_r for any particular r can be used as a test statistic, r being at our disposal. An obvious choice is the median value, i.e. $r=\frac{1}{2}(n+1)$ for n odd and $r=\frac{1}{2}n$ (or $\frac{1}{2}n+1$) for n even. Since on the alternative we would expect w_r to be smaller than on the null hypothesis a one-sided F-test is appropriate.

More important is the second of the three tests which we call the *modified Kolmogorov test*. This is obtained by considering the difference between the sample and population distribution functions corresponding to $w_1,\ldots,w_n$. The test statistic is

$$K_m = \max_{r=1,\ldots,n}\left(\frac{r}{n}-w_r\right). \tag{9}$$

Since we expect K_m to increase as we depart from the null hypothesis a one-sided test is appropriate. Consequently the test procedure is to reject when K_m is greater than the value tabulated for a one-sided Kolmogorov test. A suitable table has been given by Miller (1956).

The third test comes from the observation that $\prod_{j=1}^{n} u_{(j)} = \prod_{j=1}^{n} u_j$. Since $\prod_{j=1}^{n} w_j$ is distributed like $\prod_{j=1}^{n} u_{(j)}$ it therefore has the same distribution as $\prod_{j=1}^{n} u_j$. Karl Pearson (1933) pointed out that $\prod_{j=1}^{n} u_j$ is distributed like $\exp(-\frac{1}{2}\chi^2)$, where χ^2 has the χ^2 distribution with $2n$ degrees of freedom, and hence can be used as the statistic for an exact test of goodness-

of-fit. Applying Pearson's suggestion to $\prod_{j=1}^{n} w_j$ we obtain a third exact test which we call the *modified probability product test*. The test statistic is

$$P_m = -2 \log \prod_{j=1}^{n} w_j \tag{10}$$

which is tested as a χ^2 variate with $2n$ degrees of freedom.

The power of the original probability product test was studied by E. S. Pearson (1939) and in the light of his results it seems likely that the P_m test will in theory have high power against a wide class of alternatives. However, its use in practice in the form presented above is not recommended since the value of the statistic P_m seems to be too much affected by rounding errors in $u_1, \ldots, u_n$. It is possible that the test might be altered slightly so as to avoid this difficulty but the best way of doing so is not obvious.

Some empirical results obtained by applying the first two tests to artificial data are given in the next section.

4. SOME EMPIRICAL RESULTS

The performance of the M_r and K_m tests proposed in § 3 has been compared with the performance of the χ^2 and ordinary Kolmogorov tests on five samples of 50 observations from each of the following three distributions,

$$\textit{Exponential:} \quad \exp\{-(x+1)\}\,dx \qquad (x \geqslant -1). \tag{11}$$

$$\textit{Laplace:} \quad 2^{-\frac{3}{2}} \exp(-2^{-\frac{1}{2}}|x|)\,dx \qquad (-\infty \leqslant x \leqslant \infty). \tag{12}$$

$$\textit{Normal:} \quad (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2)\,dx \qquad (-\infty \leqslant x \leqslant \infty). \tag{13}$$

Each distribution has mean zero and variance unity and the null hypothesis is that the true distribution is the $N(0, 1)$ distribution defined by (13). Random deviates from these distributions have been tabulated by Quenouille (1959) and the five sets of three samples in this experiment are the values given on the first five pages of Quenouille's tables. Deviates for distributions (11) and (12) were obtained by Quenouille by transformation of the corresponding deviates from distribution (13); consequently the samples from the three distributions correspond in the sense of being composed of values with the same probability integral transforms.

The following four tests were applied to each sample of 50 observations:

(*a*) The χ^2 test, taking ten groups with an expected number of five in each group.

(*b*) The two-sided Kolmogorov test based on the statistic

$$K = \max_{j=1,\ldots,n} |S(x_j) - F(x_j)|,$$

where $S(x)$ is the sample distribution function and $F(x)$ is the hypothetical distribution function. Significance was assessed by referring to the tables of critical values of this statistic published by Miller (1956).

(*c*) The M_r test proposed in § 3 taking $r = 25$.

(*d*) The modified Kolmogorov test based on the statistic K_m defined by (9).

The results are given in Table 1. The entries are values of the appropriate test statistics. Significance at the 5 and 1% levels are denoted by single and double asterisks, respectively.

With due reservations owing to the small number of samples considered, the results suggest the following tentative conclusions for the sample size and alternatives considered:

(i) The unmodified Kolmogorov test is less powerful than the χ^2 test.

(ii) The modified median test is more powerful than the unmodified Kolmogorov test but slightly less powerful than the χ^2 test.

(iii) The modified Kolmogorov test is more powerful than the unmodified Kolmogorov test, slightly more powerful than the modified median test and about as powerful as the χ^2 test.

Table 1. *Comparison of performances of new and existing tests on artificial samples*

		Sample				
Population	Test	1	2	3	4	5
Exponential	χ^2	17·2*	24·0**	22·8**	38·0**	24·8**
	K	0·16	0·18	0·18	0·26**	0·21**
	M_{25}	2·20**	1·71*	2·48**	2·44**	1·61*
	K_m	0·23**	0·22**	0·25**	0·22**	0·18*
Laplace	χ^2	26·8**	13·5	12·0	19·6*	18·4*
	K	0·19*	0·15	0·10	0·14	0·14
	M_{25}	2·38**	1·23	1·19	0·91	1·36
	K_m	0·25**	0·17*	0·08	0·17*	0·16
Normal	χ^2	7·2	13·7	8·8	7·2	8·8
	K	0·17	0·10	0·15	0·14	0·10
	M_{25}	1·62*	1·24	0·93	0·70	1·21
	K_m	0·16	0·12	0·05	0·12	0·08

The critical values of the four statistics are:

	χ^2	K	M_{25}	K_m
5%	16·92	0·188	1·609	0·170
1%	21·67	0·226	1·967	0·211

In making these comparisons with the χ^2 test it should be remembered that the χ^2 test is not an exact test whereas the other three tests are exact. There is no suggestion that the χ^2 test should be used as an absolute standard; it is merely a convenient yardstick which is used because of familiarity among statisticians. Moreover, I would like to make it clear that in presenting these results it has not been my intention to make any general statements regarding the relative powers of the various procedures; the intention has been merely to summarize the performance of the tests under the conditions of this experiment.

5. Applications to time-series and arrival-time distributions

A further application of the foregoing theory is to testing the hypothesis that a series of identically distributed normal variables have uniform spectral density, i.e. are serially independent. In the spectral approach to time-series the basic statistic is the periodogram

$$p_j = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t \exp\left[(2\pi ijt)/T\right]\right|^2 \quad (j = 0, 1, \ldots, \tfrac{1}{2}T \text{ for } T \text{ even}),$$

$$(j = 0, 1, \ldots, \tfrac{1}{2}(T-1) \text{ for } T \text{ odd}). \qquad (14)$$

Let $m = \frac{1}{2}T - 1$ for T even and $m = \frac{1}{2}(T-1)$ for T odd. On the null hypothesis that $x_1, \ldots, x_T$ are independent $N(\mu, \sigma^2)$ variables the quantities $2\pi p_j/\sigma^2$ $(j = 1, \ldots, m)$ are independent exponential variables with density $\exp(-x)$. Putting $c_j = p_j \Big/ \sum_{i=1}^{m} p_i$ $(j = 1, \ldots, m)$ we have the well-known result that $c_1, \ldots, c_m$ are distributed like the intervals between successive ordered $U(0, 1)$ variables, i.e. they have the distribution (2) with $n + 1 = m$. Transforming to $c_{(1)}, \ldots, c_{(m)}$, then to $g_1, \ldots, g_{m-1}$ and finally to $w_1, \ldots, w_{m-1}$ as in § 2, we have reduced the problem to testing the hypothesis that $w_1, \ldots, w_{m-1}$ are ordered $U(0, 1)$ variables. For this purpose the modified Kolmogorov test recommended in § 3 is appropriate.

It might be asked what advantage this test is likely to have over existing exact tests such as the von Neumann test (1941) based on the statistic

$$\delta^2/s^2 = T \sum_{t=2}^{T} (x_t - x_{t-1})^2 \Big/ (T-1) \sum_{t=1}^{T} (x_t - \bar{x})^2,$$

Fisher's test (1929) based on the largest of the c_j's, or indeed the ordinary Kolmogorov test applied to the unordered c's. The answer is similar to that given for the goodness-of-fit problem in § 3, namely, that we expect it to have high power against a wider range of alternatives.

The method applies also in a rather obvious way to testing the hypothesis that events are occurring in a Poisson process. Suppose that one observes a Poisson process for a fixed length of time T and that n events occur at times $t_1, t_2, \ldots, t_n$ during the interval $(0, T)$. Let $u_{(j)} = t_j/T$ $(j = 1, \ldots, n)$. Then $u_{(1)}, \ldots, u_{(n)}$ are distributed in the form (1). Alternatively, suppose one fixes n in advance and starting at time zero, records the times $t_1, \ldots, t_{n+1}$ at which $n + 1$ events occur. Let $u_{(j)} = t_j/t_{n+1}$ $(j = 1, \ldots, n)$. Then $u_{(1)}, \ldots, u_{(n)}$ are distributed in the form (1). Simple proofs of these results are given by Epstein (1960, Appendix 1). In either case one can transform to the c's, the g's and the w's as in §§ 2 and 3 and employ the modified Kolmogorov test in the way described.

6. Tests based on systematic subsampling of the observations

Most of the work required to carry out the goodness-of-fit tests described in § 3 is likely to occur in making the probability integral transformation from the x's to the u's; indeed when the sample is large the amount of labour required can be prohibitively great. A way of reducing the work substantially is to confine the analysis to a systematic subsample of the observations. Suppose that the sample size is now denoted by N. Instead of transforming all the x's we first arrange them in order of magnitude, giving $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(N)}$ say, and then transform every kth, i.e. we calculate $u_{(k)} = F(x_{(k)})$, $u_{(2k)} = F(x_{(2k)}), \ldots, u_{(nk)} = F(x_{(nk)})$, where k is a suitable chosen integer and n is the largest integer satisfying $nk + k - 1 \leqslant N$.

Suppose for simplicity that $N = nk + k - 1$; otherwise reject $N - nk - k + 1$ observations at random. The distribution of $u_{(k)}, \ldots, u_{(nk)}$ is (see, for example, Wilks, (1948))

$$dP = \frac{(nk+k-1)!}{\{(k-1)!\}^{n+1}} u_{(k)}^{k-1} \prod_{j=1}^{n-1} (u_{(jk+k)} - u_{(jk)})^{k-1} (1 - u_{(nk)})^{k-1} \, du_{(k)} \ldots du_{(nk)}$$
$$(0 \leqslant u_{(k)} \leqslant u_{(2k)} \leqslant \ldots \leqslant u_{(nk)} \leqslant 1). \qquad (15)$$

Let

$$\left.\begin{aligned} p_j &= \frac{u_{(jk)}}{u_{(jk+k)}} \\ p_n &= u_{(nk)} \end{aligned}\right\} \quad (j = 1, \ldots, n-1). \qquad (16)$$

The Jacobian is $$\frac{\partial(u_{(k)}, \ldots, u_{(nk)})}{\partial(p_1, \ldots, p_n)} = \prod_{j=2}^{n} u_{(jk)}.$$

The distribution of $p_1, \ldots, p_n$ is, therefore,

$$dP = \frac{(nk+k-1)!}{\{(k-1)!\}^{n+1}} p_1^{k-1}(1-p_1)^{k-1} p_2^{2k-1}(1-p_2)^{k-1} \ldots p_n^{nk-1}(1-p_n)^{k-1}\, dp_1, \ldots, dp_n$$
$$(0 \leqslant p_j \leqslant 1, \text{ all } j). \qquad (17)$$

Thus $p_1, \ldots, p_n$ are independently distributed, the distribution of p_j being of the Beta form

$$dP = \{B(jk, k)\}^{-1} p_j^{jk-1}(1-p_j)^{k-1}\, dp_j. \qquad (18)$$

Let z_j denote the probability integral transform of p_j, i.e.

$$z_j = \{B(jk, k)\}^{-1} \int_0^{p_j} x^{jk-1}(1-x)^{k-1}\, dx \quad (j = 1, \ldots, n). \qquad (19)$$

Then $z_1, \ldots, z_n$ are independent $U(0, 1)$ variables. The actual values of the z's determined by (19) may be obtained from the *Tables of the Incomplete Beta Function* (K. Pearson, 1948).

Let
$$q_j = z_j^{1/j} \quad (j = 1, \ldots, n). \qquad (20)$$

The joint distribution of $q_1, \ldots, q_n$ is seen to be

$$dP = n!\, q_2 q_3^2 \ldots q_n^{n-1}\, dq_1 \ldots dq_n \quad (0 \leqslant q_j \leqslant 1, \text{ all } j). \qquad (21)$$

This is of the form (17) with $k = 1$. Let

$$v_j = \prod_{i=j}^{n} q_i. \qquad (22)$$

We observe that $q_j = v_j/v_{j+1}$ and $q_n = v_n$. Consequently the distribution of $v_1, \ldots, v_n$ is given by putting $k = 1$ in (15), i.e.

$$dP = n!\, dv_1 \ldots dv_n \quad (0 \leqslant v_1 \leqslant v_2 \leqslant \ldots \leqslant v_n \leqslant 1). \qquad (23)$$

It follows that $v_1, \ldots, v_n$ are distributed as ordered $U(0, 1)$ variables.

We have transformed the systematic subsample obtained by taking every kth value into a sample of ordered $U(0, 1)$ variables and may therefore now apply to $v_1, \ldots, v_n$ the methods developed in §§ 2 and 3 for analysing $u_{(1)}, \ldots, u_{(n)}$. Of course a substantial amount of computation is required in order to proceed from $u_{(k)}, \ldots, u_{(nk)}$ to $v_1, \ldots, v_n$. Whether or not this is worth while depends on how much computation is saved by analysing a systematic subsample instead of the whole sample. The procedure clearly involves some loss of power but I have not attempted to assess how great this is likely to be. Note that the same method can be employed to construct an exact test for equality of variances; however, in view of the amount of computation required the idea does not seem worth elaborating upon in detail.

7. The elimination of nuisance parameters by the method of random substitution

A serious limitation to the practical usefulness of the Kolmogorov test has been that there has not been available a simple method of allowing for the presence of nuisance parameters. For example, it has not been possible to use it to obtain an exact test of normality in the usual situation where the population mean and variance are unknown. The reader is

referred to the paper by Kac, Kiefer & Wolfowitz (1955) for a thorough discussion of the difficulties involved. A similar limitation applies to the tests proposed in § 3. Consequently, if we are to bring these tests to the point of practical usefulness for the classical goodness-of-fit problem we require a practicable method of dealing with nuisance parameters.

A simple randomization method for doing this will now be presented. Although the method is of wide application it is convenient to introduce it by applying it to the specific problem of testing for normality.

Suppose we wish to test the hypothesis that $x_1, \ldots, x_n$ are independent observations from a normal distribution with unknown mean μ and unknown variance σ^2. Let $\bar{x} = n^{-1}\Sigma x_i$ and $s^2 = (n-1)^{-1}\Sigma(x_i - \bar{x})^2$; furthermore let $\bar{x}'$, s'^2 be observations of random variables independent of $x_1, \ldots, x_n$ and distributed as the sample mean and variance calculated from a sample of n independent observations from a $N(0, 1)$ distribution, i.e. $\bar{x}'$ has the distribution

$$dP = \text{constant} \times \exp\left(-\tfrac{1}{2}n\bar{x}'^2\right) d\bar{x}' \tag{24}$$

and s'^2 has the distribution

$$dP = \text{constant} \times (s'^2)^{(\frac{1}{2}n-\frac{3}{2})} \exp\left\{-\tfrac{1}{2}(n-1)s'^2\right\} ds'^2. \tag{25}$$

Define $x_1', \ldots, x_n'$ by the relations

$$\frac{x_i' - \bar{x}'}{s'} = \frac{x_i - \bar{x}}{s} \quad (i = 1, \ldots, n). \tag{26}$$

We shall show that $x_1', \ldots, x_n'$ are independent $N(0, 1)$ variables. Any of the exact tests previously discussed can now be applied to $x_1', \ldots, x_n'$. We have, in fact, transformed the composite hypothesis concerning $x_1, \ldots, x_n$ into a simple hypothesis concerning $x_1', \ldots, x_n'$.

The proof is as follows. For the distribution of $x_1, \ldots, x_n$ we have

$$dP = \text{constant} \times \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i - u)^2\right\} dx_1 \ldots dx_n. \tag{27}$$

Let

$$l_i = \frac{x_i - \bar{x}}{s} \qquad (i = 1, \ldots, n), \tag{28}$$

$$m_r = \left(\frac{r-1}{r}\right)^{\frac{1}{2}}\left(l_r - \frac{1}{r-1}\sum_{i=1}^{r-1} l_i\right) \quad (r = 2, \ldots, n); \tag{29}$$

$$\left.\begin{aligned} m_2 &= (n-1)^{\frac{1}{2}} \cos a_1 \cos a_2 \ldots \cos a_{n-2}, \\ m_r &= (n-1)^{\frac{1}{2}} \cos a_1 \cos a_2 \ldots \cos a_{n-r} \sin a_{n-r+1} \quad (r = 3, \ldots, n-1), \\ m_n &= (n-1)^{\frac{1}{2}} \sin a_1. \end{aligned}\right\} \tag{30}$$

(29) and (30) are the Helmert and polar transformations. Substituting in (27) we have the classical representation of the probability element of a normal sample in terms of $\bar{x}$, s^2 and the angular variables $a_1, \ldots, a_{n-2}$, i.e.

$$\begin{aligned} dP = \text{constant} \times \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right\} s^{n-3} \exp\left\{-\tfrac{1}{2}(n-1)s^2/\sigma^2\right\} \\ \times \cos^{n-3} a_1 \cos^{n-4} a_2 \ldots \cos a_{n-3}\, d\bar{x}\, ds^2\, da_1 \ldots da_{n-2}. \end{aligned} \tag{31}$$

(Geary, 1933; see also Kendall & Stuart, 1958, pp. 247, 250 for details of the derivation).

Consequently $\bar{x}$, s^2 and the set of angles $a_1, \ldots, a_{n-2}$ are independently distributed with distributions

$$dP = \text{constant} \times \exp\left\{\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} d\bar{x}, \tag{32}$$

$$dP = \text{constant} \times s^{n-3} \exp\{-\tfrac{1}{2}(n-1)\, s^2/\sigma^2\} ds^2, \tag{33}$$

$$dP = \text{constant} \times \cos^{n-3} a_1 \cos^{n-4} a_2 \ldots \cos a_{n-3}\, da_1 \ldots da_{n-2}, \tag{34}$$

respectively.

More important from the present point of view is the converse result which for the sake of clarity we state in the form of the following lemma.

LEMMA. Suppose $\bar{x}, s^2, a_1 \ldots a_{n-2}$ are independent random variables with distributions (32), (33) and (34). Let $x_1, \ldots, x_n$ be determined by (28), (29) and (30) together with the relation $\sum_{i=1}^{n} l_i = 0$. Then $x_1, \ldots, x_n$ are independent $N(\mu, \sigma^2)$ variables.

To prove this, suppose that s^2 and $a_1, \ldots, a_{n-2}$ are independent with distributions (33) and '34), respectively. Let

$$\left.\begin{aligned} y_2 &= (n-1)^{\frac{1}{2}}\, s \cos a_1 \cos a_2 \ldots \cos a_{n-2}, \\ y_r &= (n-1)^{\frac{1}{2}}\, s \cos a_1 \cos a_2 \ldots \cos a_{n-r} \sin a_{n-r+1} \quad (r = 3, \ldots, n-1), \\ y_n &= (n-1)^{\frac{1}{2}}\, s \sin a_1. \end{aligned}\right\} \tag{35}$$

The Jacobian is

$$\frac{\partial(y_2, \ldots, y_n)}{\partial(s^2, a_1, \ldots, a_{n-2})} = \tfrac{1}{2}(n-1)^{\frac{1}{2}(n-1)} s^{n-3} \cos^{n-3} a_1 \cos^{n-4} a_2 \ldots \cos a_{n-3}$$

(Kendall & Stuart, 1958, p. 247). On substitution we find for the distribution of $y_2, \ldots, y_n$

$$dP = \text{constant} \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=2}^{n} y_i^2\right) dy_2 \ldots dy_n.$$

Suppose $\bar{x}$ independently has distribution (32). Let $x_1, \ldots, x_n$ be determined by

$$\sum_{i=1}^{n} x_i = n\bar{x}, \tag{36}$$

$$\left(\frac{r-1}{r}\right)^{\frac{1}{2}} \left(x_r - \frac{1}{r-1} \sum_{i=1}^{r-1} x_i\right) = y_r \quad (r = 2, \ldots, n). \tag{37}$$

On substituting in the joint distribution of $\bar{x}$ and $y_2, \ldots, y_n$ we find that $x_1, \ldots, x_n$ have the distribution (27). It remains to show that (35), (36) and (37) are equivalent to (28), (29), (30) and the relation $\sum_{i=1}^{n} l_i = 0$; the equivalence follows on summing both sides of (28) and on multiplying (30) through by s.

Suppose now that $x_1, \ldots, x_n$ have distribution (27) and that $a_1, \ldots, a_{n-2}$ are determined by (28), (29) and (30). Let $\bar{x}'$ and s'^2 have distributions (24) and (25) independently of $x_1, \ldots, x_n$. Suppose that $l_1, \ldots, l_n$ are determined from $a_1, \ldots, a_{n-1}$ by (29), (30) and the relation $\sum_{i=1}^{n} l_i = 0$. Let $x_1', \ldots, x_n'$ be determined by

$$\frac{x_i' - \bar{x}'}{s'} = l_i \quad (i = 1, \ldots, n). \tag{38}$$

By the lemma it follows that $x'_1, \ldots, x'_n$ are independent $N(0, 1)$ variables. We have therefore established the claim that composite hypotheses concerning $x_1, \ldots, x_n$ can be tested as simple hypotheses concerning $x'_1, \ldots, x'_n$.

I want to put on record the fact that my original geometrical derivation of this result was much shorter than that presented here. However, a number of my friends were so sceptical about either the result itself or the argument supporting it that I felt compelled to supply full algebraical details, even though they might seem rather tedious (see also Prof. Pearson's note on p. 55).

The method is clearly of wide application. Suppose in general that $x_1, \ldots, x_n$ have a distribution with distribution function $F(x_1, \ldots, x_n, \theta)$ depending on a set of parameters θ for which a sufficient set of statistics t_1 is available. Suppose that a transformation T independent of θ carrying $x_1, \ldots, x_n$ into (t_1, t_2) exists, t_2 being another set of statistics distributed independently of t_1, such that the inverse transformation T^{-1} carries (t_1, t_2) into $x_1, \ldots, x_n$ uniquely. Let $G(t_1, \theta)$ denote the distribution function of t_1 and let t'_1 be an observation of a random vector independent of $x_1, \ldots, x_n$ with distribution function $G(t'_1, \theta_0)$ where θ_0 is a known value of θ. Let $x'_1, \ldots, x'_n$ be the values obtained by applying T^{-1} to (t'_1, t_2). Then $x'_1, \ldots, x'_n$ are distributed with distribution function $F(x'_1, \ldots, x'_n, \theta_0)$. As for the lemma, the proof follows by merely reversing the transformation.

To illustrate the use of this general result consider the construction of an exact distribution-free test of serial correlation. The null hypothesis is that $x_1, \ldots, x_n$ are independently and identically distributed with an unknown continuous distribution function. Let $x_{i_1}, x_{i_2}, \ldots, x_{i_n}$ be the set of x's arranged in increasing order of magnitude; $i_1, i_2, \ldots, i_n$ is then a permutation of the suffixes $1, 2, \ldots, n$. Let $x'_{i_1}, x'_{i_2}, \ldots, x'_{i_n}$ be a sample of n independent $N(0, 1)$ variables taken from a table of random normal deviates and arranged in increasing order of magnitude. The arrangement of the suffixes is taken to be the same as on $x_{i_1}, x_{i_2}, \ldots, x_{i_n}$. Calculate von Neumann's statistic

$$\frac{\delta^2}{s^2} = \frac{n \sum_{t=2}^{n} (x'_t - x'_{t-1})^2}{(n-1) \sum_{t=1}^{n} (x'_t - \bar{x}')^2},$$

where

$$\bar{x}' = \frac{1}{n} \sum_{t=1}^{n} x'_t$$

and refer to tables of critical values of this statistic (Hart, 1942)

The key to the construction of this test is the observation that the order statistic $x_{i_1}, x_{i_2}, \ldots, x_{i_n}$ is a sufficient statistic for the unknown distribution function which, when the null hypothesis is true, is distributed independently of the permutation $i_1, \ldots, i_n$ of the suffixes $1, 2, \ldots, n$ (see Fraser, 1957). Other possible applications which spring to mind are the testing of bivariate and multivariate normality and the testing of serial correlation in least-squares regression.

The price that has to be paid for the elimination of nuisance parameters by this method is that an element of randomization is introduced in the analysis of the data. The device can be objected to on the ground that it permits different investigators to draw different conclusions from the same set of data. The reader is referred to the paper by E. S. Pearson (1950) for a searching discussion of the questions this raises. Some statisticians object to the use of randomization at any stage, design or analysis, as a matter of principle; others feel that randomization is legitimate in the design of an experiment but not in the analysis of the

results. I do not wish to take up space on this occasion by discussing my personal attitude to randomization. I do, however, wish to point out an important operational distinction between randomization in design and randomization in analysis. In design an act of randomization is performed once and for all, the experimental layout being arranged according to what turns up in the randomization procedure. Since the experimental results achieved depend on the result of randomization, it is hard for the investigator to see how randomization has affected his conclusions without repeating the entire experiment. And even if he were to do this, any differences which emerged could just as well be due to changes in other conditions of the experiment as to a change in the result of the randomization procedure.

Table 2. *Some results for random-substitution tests of normality*

		Random pair $\bar{x}'$, s'^2				
Population	Test	1	2	3	4	5
Exponential	χ^2	21·2*	26·0**	21·6*	34·4**	22·8**
	K	0·160	0·227**	0·271**	0·288**	0·181
	M_{25}	2·44**	2·22**	2·23**	2·61**	2·43**
	K_m	0·263**	0·274**	0·278**	0·295**	0·294**
Laplace	χ^2	12·0	25·2**	14·8	22·0**	22·4**
	K	0·163	0·190*	0·191*	0·256**	0·156
	M_{25}	2·42**	2·66**	2·12**	2·72**	2·72**
	K_m	0·222**	0·242**	0·209*	0·240*	0·239**
Normal	χ^2	8·4	6·0	13·6	11·6	8·8
	K	0·126	0·138	0·190	0·100	0·120
	M_{25}	1·75*	1·45	1·76*	1·45	1·39
	K_m	0·129	0·136	0·147	0·143	0·143

At the analysis stage, however, the position is different. The experimental results are a fixed set of numbers and the investigator can repeat the randomization procedure as often as he likes without affecting them. Consequently if he wishes to know the extent to which his conclusions are affected by the result of randomization he merely has to repeat the analysis after a fresh act of randomization. This can be done as often as the investigator wishes. Of course, in order to preserve the exact probabilities of his test he must abide by the results indicated by his first act of randomization.

As an illustration, the first sample of 50 for each of the three distributions considered in § 4 was tested for normality assuming the population mean and variance to be unknown. The sample means and variances were replaced by values $\bar{x}'$, s'^2 picked at random. For $\bar{x}'$ the quantity $1/\sqrt{50}$ times a random $N(0, 1)$ deviate was used and for s'^2 the quantity

$$(1/49)\left(2\sum_{j=1}^{24} e_j + z^2 + 48\right)$$

was used, where $e_1, \ldots, e_{24}$ are random exponential deviates with distribution (11) while z is a random $N(0, 1)$ deviate; all these random deviates were taken from Quenouille's tables. The hypothesis that the values $x'_1, \ldots, x'_{50}$ determined by (38) were independent $N(0, 1)$ variables was tested by each of the four tests considered in § 4. In order to explore the effect of randomization the tests were repeated five times using a new pair of randomly chosen values $\bar{x}$, s'^2 on each occasion. The results are given in Table 2. Entries in the table are values of the appropriate test statistics. Significance at the 5 and 1 % levels are denoted by single and double asterisks, respectively.

It has to be admitted that the amount of variation from random pair to random pair is disappointing for both the χ^2 test and the ordinary Kolmogorov test. The reason is presumably the strong dependence of the test statistics on the sample mean and variance. On the other hand the amount of variation for the modified median and modified Kolmogorov tests is substantially less, indicating that the effect of the re-ordering of the intervals has been to reduce the dependence of the test statistics on the randomly substituted values.

I am indebted to D. E. Barton, M. G. Kendall and A. Stuart for reading the manuscript and for some helpful comments.

REFERENCES

BARNARD, G. A. (1953). Time intervals between accidents—a note on Maquire, Pearson and Wynn's paper. *Biometrika*, **40**, 212–3.

BARTHOLEMEW, D. J. (1956). Tests for randomness in a series of events when the alternative is a trend. *J. R. Statist. Soc.* B, **18**, 234–239.

BARTHOLEMEW, D. J. (1957). Testing for departure from the exponential distribution. *Biometrika*, **44**, 253–6.

BARTON, D. E. & DAVID, F. N. (1955). Sums of ordered intervals and distances. *Mathematika*, **2**, 150–9.

BARTON, D. E. & DAVID, F. N. (1956*a*). Tests for randomness of points on a line. *Biometrika*, **43**, 104–12.

BARTON, D. E. & DAVID, F. N. (1956*b*). Some notes on ordered random intervals. *J. R. Statist. Soc.* B, **18**, 79–94.

COX, D. R. (1955). Some statistical methods connected with series of events. *J. R. Statist. Soc.* B, **17**, 129–57.

DARLING, D. A. (1953). On a class of problems relating to the random division of an interval. *Ann. Math. Statist.* **24**, 239–53.

DARLING, D. A. (1957). The Kolmogorov–Smirnov, Cramer–von Mises tests. *Ann. Math. Statist.* **28**, 823–38.

DWASS, M. (1959). The distribution of linear combinations of random divisions of an interval. O.N.R. Research Memorandum No. 21 (unpublished). Systems Research group, Northwestern University.

EPSTEIN, B. (1960). Tests for the validity of the assumption that the underlying distribution of life is exponential, Part I. *Technometrics*, **2**, 83–101.

EPSTEIN, B. & SOBEL, M. (1954). Some theorems relevant to life-testing from an exponential distribution. *J. Amer. Statist. Ass.* **25**, 373–81.

EPSTEIN, B. & TSAO, C. K. (1953). Some two-sample tests based on ordered observations from the exponential distribution. *Ann. Math. Statist.* **24**, 458–66.

FISHER, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc.* A, **125**, 54–9.

FRASER, D. A. S. (1957). *Non-parametric Methods in Statistics*. New York: Wiley.

GEARY, R. C. (1933). A general expression for the moments of certain symmetrical functions of normal samples. *Biometrika*, **25**, 184–6.

GREENWOOD, M. (1946). The statistical study of infectious diseases. *J. R. Statist. Soc.* A, **109**, 85–103.

HART, B. I. (1942). Significance levels for the ratio of the mean-square successive difference to the variance. *Ann. Math. Statist.* **13**, 445–7.

IRWIN, J. O. (1955). A unified derivation of some well-known frequency distributions of interest in biometry and statistics. *J. R. Statist. Soc.* A, **118**, 389–98.

KAC, M., KIEFER, J. & WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* **26**, 189–211.

KENDALL, M. G. & STUART, A. (1958). *The Advanced Theory of Statistics*, vol. I. London: Griffin.

MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika*, **39**, 168–80.

MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1953). Further notes on the analysis of accident data. *Biometrika*, **40**, 213–16.

MALMQUIST, S. (1951). On a property of order statistics from a rectangular population. *Skand. Aktuar. Tidskr.* **33**, 214–22.

MAULDON, J. G. (1951). Random division of an interval. *Proc. Camb. Phil. Soc.* **47**, 331–6.

MILLER, L. H. (1956). Tables of percentage points of Kolmogorov statistics. *J. Amer. Statist. Ass.* **51**, 111–21.
MORAN, P. A. P. (1947). The random division of an interval, I. *J. R. Statist. Soc.* B, **9**, 92–8.
MORAN, P. A. P. (1951). The random division of an interval, II. *J. R. Statist. Soc.* B, **13**, 147–50.
NEYMAN, J. (1937). 'Smooth test' for goodness of fit. *Skand. Aktuar Tidskr.* **20**, 149–99.
PEARSON, K. (1948). (Ed.) *Tables of the Incomplete Beta Function.* London: Biometrika Office.
PEARSON, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, **25**, 379–410.
PEARSON, E. S. (1939). The probability integral transformation for testing goodness-of-fit and combining independent tests of significance. *Biometrika*, **30**, 134–48.
PEARSON, E. S. (1950). On questions raised by the combination of tests based on discontinuous distributions. *Biometrika*, **37**, 383–98.
QUENOUILLE, M. H. (1959). Tables of random observations from standard distributions. *Biometrika*, **46**, 178–202.
RENYI, A. (1953). On the theory of order statistics. *Acta math. hung.* **4**, 191–231.
SHERMAN, B. (1950). A random variable related to the spacing of sample values. *Ann. Math. Statist.* **21**, 339–61.
SUKHATME, P. V. (1936). On the analysis of k samples from exponential populations with special reference to problems of random intervals. *Statist. Res. Mem.* **1**, 94–112.
SUKHATME, P. V. (1937). Tests of significance for samples of the χ^2 population with two degrees of freedom. *Ann. Eugen., Lond.*, **8**, 52–6.
WHITWORTH, W. A. (1887). *Choice and Chance*, 3rd ed. Cambridge University Press.
WILKS, S. S. (1948). Order statistics. *Bull. Amer. Math. Soc.* **54**, 6–50.
VAN DANTZIG, D. (1954). Mathematical problems raised by the flood disaster 1953. *Proceedings of the International Congress of Mathematicians*, 1954, vol. I, 218–39. Amsterdam.
VON NEUMANN, J. (1941). Distribution of the ratio of the mean-square successive difference to the variance. *Ann. Math. Statist.*, **12**, 367–95.

EDITORIAL NOTE

If we write $u_i = (x_i - \bar{x})/s$, Mr Durbin's equation (26) may be written

$$x_i' = \bar{x}' + s'u_i, \tag{1}$$

where $\bar{x}'$, s' and u_i are three independent random variables, the first distributed normally with variance $1/n$ and the second as $\chi_{n-1}/\sqrt{(n-1)}$. The n values of u_i are of course not independent, but if x_i is a random member of the original sample it is known (W. R. Thompson, 1935, *Ann. Math. Statist.* **6**, 214 and E. S. Pearson & C. Chandra Sekar, 1936, *Biometrika*, **28**, 308) that the p.d.f. of u_i is

$$p(u_i) = \text{constant} \times \left(1 - \frac{n}{(n-1)^2}u_i^2\right)^{\frac{1}{2}(n-4)}, \quad -\frac{n-1}{\sqrt{n}} \leqslant u_i \leqslant \frac{n-1}{\sqrt{n}}. \tag{2}$$

At first sight it might appear unlikely that the product, $s'u_i$ of independent variables following, respectively, a χ and a symmetrical β distribution would be normally distributed, as they must be if x_i' is to be normally distributed. However, on performing the necessary transformation and integration it is found that $s'u_i$ *is* distributed normally about zero with variance $(n-1)/n$. It follows at once that x_i' is an $N(0, 1)$ variable.

Mr Durbin's proof in terms of random angles is clearly the more fundamental, but he has suggested that I should add this note since several readers of his paper in typescript had been puzzled, as I was, about the distribution of $s'u_i$.

E. S. P

Biometrika (1961), **48**, 1 *and* 2, *p.* 57
Printed in Great Britain

Preemptive priority queueing

BY C. R. HEATHCOTE
Australian National University

SUMMARY

The basic queueing model considered is one in which certain customers are given a preemptive right to service over routine, non-priority, customers. Servicing of the latter is thus liable to interruption by the arrival of a priority customer. Given a model of this kind we require the distribution of queue length for the non-priority class of customers. The service time distribution of the non-priority customers is assumed to follow χ^2 with an even number of degrees of freedom. This includes the negative exponential and constant service time distributions as special cases. The method of generating functions is used to obtain a solution in the equilibrium state and, as a Laplace transform, for finite time. Breakdowns in the service mechanism are interpreted as a modification of the preemptive discipline.

1. INTRODUCTION

A problem of interest in the practical application of the theory of queues is that of the effect on queue behaviour of interruptions to the servicing of customers. One such case is when breakdowns in the service mechanism add to the usual delay. Alternatively the interruptions may be caused by a queue discipline which assigns priority to a certain group of customers. For example, in a communication system a priority customer could be an urgent message to be transmitted immediately on arrival irrespective of the state of the queue of routine (non-priority) messages. Under this discipline the arrival of a priority customer and a breakdown in the service mechanism are equivalent from the point of view of the customer whose service has been interrupted. A model combining both priority discipline and breakdowns can be postulated, in which case a breakdown is formally interpreted as a priority customer with precedence over all others. We are interested in finding to what extent interruptions such as these affect the distribution of queue length.

Queueing models in which service interruptions are a distinguishing feature are termed preemptive priority systems in distinction to the priority discipline in which a priority customer proceeds to the head of the waiting line on arrival, but waits until the service of the current customer has ended. The preemptive priority model may be described as follows. A service facility caters for a population of customers divided into R priority classes, $R = 2, 3, \ldots$. (The case $R = 1$ corresponds to no priorities.) Label the classes serially in order of precedence $1, 2, \ldots, R$. Then on arrival a customer of class i commences service immediately provided no members of classes $1, 2, \ldots, i$, are present. If customers belonging to any of these classes are present then the new arrival joins the queue of members of the same priority class. The servicing of customers of class i does not commence until the system is empty of customers of classes $1, 2, \ldots, i-1$. Further, a rule must be given governing the manner in which the service of an interrupted customer is to be resumed when the interruption ceases. These are the assumptions specific to preemptive queueing. We also assume that the queue discipline within each class is 'first come, first served', and that all interarrival and service times are independently distributed. The inclusion of breakdowns in this

description implies that times between their occurrence and repair times are equivalent to inter-arrival and service times, respectively.

Preemptive queueing was apparently first considered by Barry (1956), and in more detail by White & Christie (1958) and Stephan (1958). These authors discussed the steady-state model for a single server negative exponential queue with two priority classes. Heathcote (1959) obtained the joint generating function of the temporal probabilities for this queue. Subsequently these results were generalized to the case of R priority classes (Heathcote, 1960), but the assumption of negative exponential inter-arrival and service times was retained. In this paper the preemptive priority problem treated is that for a single server with two priority classes where the service time of the non-priority customers follows an Erlang distribution (χ^2 with an even number of degrees of freedom). A particular case considered is that when this service time is of constant duration. We assume that the inter-arrival times and the service time of priority customers follow negative exponential distributions. Models with more general service and inter-arrival distributions can be solved by the methods used here, at least in principle, but the above model is probably general enough for most practical purposes.

The question of the rule under which a customer recommences service after interruption has been discussed by White & Christie (1958). They draw a distinction between 'preemptive resume' and 'preemptive repeat'. By the former is meant that service recommences at the point it had reached when the interruption occurred. In contrast the 'preemptive repeat' rule means that service commences from the beginning every time the customer re-enters service. Provided the service distributions are negative exponential there is no difference mathematically between these two rules because of the special properties of the negative exponential distribution. Since we are concerned with a distribution which does not in general possess these special properties, we will assume that the 'preemptive resume' rule holds. The equations governing the 'preemptive repeat' rule are rather more complicated.

The preemptive priority model may be considered mathematically as a multidimensional random walk of a restricted nature. For the two-dimensional case discussed here the state of the system at time t is described by the vector (r, n); $r, n = 0, 1, 2, \ldots$, where r and n denote respectively the number of priority and non-priority class members present in the system. The co-ordinates vary discretely as customers arrive or depart, and the random walk takes place on the lattice points of the positive quadrant in the (r, n)-plane. The restricted nature of the random walk is due to the fact that if $r > 0$, n can only increase since the service mechanism has been preempted by the priority class. It is this special feature of the problem that permits a solution to be found by the straightforward use of generating functions. Further restrictions natural to the process have to be imposed. For the queueing problem itself, reflecting barriers are imposed along the r and n axes. If first passage times were of interest, such as the duration of busy periods, then these barriers would have to be modified to include an absorbing barrier at the point of interest. The breakdown model of §3 is an example where, in addition, other boundary conditions are required.

2. PREEMPTIVE PRIORITY MODEL

It is well known that the Erlang distribution may be considered as the sum of a number, K, say, of negative exponential phases, and many problems involving this distribution are solved in terms of phases, not customers, although it is the latter that are of major interest.

The equations governing transitions between phases are much simpler than those for numbers of customers, but because of its interest, we solve the problem in terms of the latter.

Let the service times of the priority and non-priority customers have probability densities

$$dH_1(t) = e^{-\mu_1 t}\mu_1 dt,$$

$$dH_2(t) = e^{-\mu_2 t}\frac{(\mu_2 t)^{K-1}}{(K-1)!}\mu_2 dt,$$

respectively. To write down the differential-difference equations governing the system the service phase of the current customer must be specified and we therefore are concerned with three discrete variables. Let $P_{rmn}(t)$ be the probability of r priority and n non-priority customers at time t, the customer in service being in phase m ($m = K, K-1, \ldots, 2, 1$). The phases are numbered in reverse order, a customer passing first through phase K, then phase $K-1$, and finally phase 1 before leaving the system. The suffix m appears only when $n > 0$ and $P_{r0}(t)$ denotes the probability of r priority and zero non-priority customers. We will assume that the system is empty at time zero, i.e.

$$P_{rmn}(0) = 0; \quad P_{r0}(0) = \delta_{r0}. \tag{2·1}$$

The method carries over directly for more general initial conditions.

The 'forward' differential-difference equations for $P_{rmn}(t)$ can be written down by considering all possible transitions from the point (r, m, n) in the time interval $(t, t+\delta t)$. In fact Morse (1958, p. 72) has given the difference equations satisfied by the equilibrium probabilities for the Poisson input, Erlang service time queue and this system can be easily generalized to incorporate the behaviour of the priority queue. Introducing Laplace transforms

$$g_{rmn}(s) = \mathscr{L}[P_{rmn}(t)] = \int_0^\infty e^{-st}P_{rmn}(t)\,dt$$

and noting that

$$\mathscr{L}\left[\frac{d}{dt}P_{rmn}(t)\right] = sg_{rmn}(s) - P_{rmn}(0),$$

it is apparent that the pure difference equations satisfied by $g_{rmn}(s)$ are almost identical with those satisfied by the equilibrium probabilities. The only equation substantially changed is that for the initial state. Writing $\alpha = s+\lambda_1+\lambda_2$, the equations for $g_{rmn}(s)$ with initial conditions (2·1) are

$$-\alpha g_{00} + \mu_1 g_{10} + \mu_2 g_{011} = -1 \quad (r = 0,\ n = 0), \tag{2·2}$$

$$-(\alpha+\mu_2)g_{0m1} + \mu_1 g_{1m1} + \mu_2 g_{0m+11} = 0 \quad (r = 0,\ n = 1,\ 1 \leqslant m < K), \tag{2·3}$$

$$-(\alpha+\mu_2)g_{0K1} + \lambda_2 g_{00} + \mu_1 g_{1K1} + \mu_2 g_{012} = 0 \quad (r = 0,\ n = 1,\ m = K), \tag{2·4}$$

$$-(\alpha+\mu_2)g_{0mn} + \lambda_2 g_{0mn-1} + \mu_1 g_{1mn} + \mu_2 g_{0m+1n} = 0 \quad (r = 0,\ n > 1,\ 1 \leqslant m < K), \tag{2·5}$$

$$-(\alpha+\mu_2)g_{0Kn} + \lambda_2 g_{0Kn-1} + \mu_1 g_{1Kn} + \mu_2 g_{01n+1} = 0 \quad (r = 0,\ n > 1,\ m = K), \tag{2·6}$$

$$-(\alpha+\mu_1)g_{r0} + \lambda_1 g_{r-10} + \mu_1 g_{r+10} = 0 \quad (r > 0,\ n = 0), \tag{2·7}$$

$$-(\alpha+\mu_1)g_{rm1} + \lambda_1 g_{r-1m1} + \mu_1 g_{r+1m1} = 0 \quad (r > 0,\ n = 1,\ 1 \leqslant m < K), \tag{2·8}$$

$$-(\alpha+\mu_1)g_{rK1} + \lambda_1 g_{r-1K1} + \lambda_2 g_{r0} + \mu_1 g_{r+1K1} = 0 \quad (r > 0,\ n = 1,\ m = K), \tag{2·9}$$

$$-(\alpha+\mu_1)g_{rmn} + \lambda_1 g_{r-1mn} + \lambda_2 g_{rmn-1} + \mu_1 g_{r+1mn} = 0 \quad (r > 0,\ n > 1,\ 1 \leqslant m \leqslant K). \tag{2·10}$$

We seek the joint generating function

$$F(s;z,y,x) = \sum_{r=0}^{\infty} z^r g_{r0}(s) + \sum_{r=0}^{\infty} z^r H_r(s;y,x), \tag{2·11}$$

where

$$H_r(s;y,x) = \sum_{m=1}^{K} \sum_{n=1}^{\infty} y^m x^n g_{rmn}(s).$$

The first sum on the right-hand side of (2·11) can be evaluated immediately. From (2·7)

$$g_{r0}(s) = g_{00}(s)\, v_1^r(s), \tag{2·12}$$

where

$$v_1(s) = (2\mu_1)^{-1}\{\alpha+\mu_1-\sqrt{[(\alpha+\mu_1)^2-4\lambda_1\mu_1]}\}.$$

Then

$$\sum_{r=0}^{\infty} z^r g_{r0}(s) = g_{00}(s)\,[1-zv_1(s)]^{-1}. \tag{2·13}$$

Multiplying (2·3)–(2·6), (2·8)–(2·10) by the appropriate power of $y^m x^n$, summing, and using (2·12), yields for $H_r(s;y,x)$

$$\mu_1 H_{r+1} - (\alpha+\mu_1-\lambda_2 x)\, H_r + \lambda_1 H_{r-1} = -\lambda_2 x y^K v_1^r g_{00} \quad (r>0), \tag{2·14}$$

$$\mu_1 H_1 - (\alpha+\mu_2-\lambda_2 x-\mu_2 y^{-1})\, H_0 = \mu_2(1-y^K x^{-1}) \sum_{n=1}^{\infty} x^n g_{01n} + y^K\{(\alpha-\lambda_2 z-\mu_1 v_1)\, g_{00}-1\} \quad (r=0). \tag{2·15}$$

The general solution of (2·14) is

$$H_r = A w_1^r + B w_2^r - y^K v_1^r g_{00},$$

where A, B are constants and w_1, w_2 are the roots of the characteristic equation, namely

$$\left.\begin{aligned} w_1(s,x) &= (2\mu_1)^{-1}\{\alpha+\mu_1-\lambda_2 x-\sqrt{[(\alpha+\mu_1-\lambda_2 x)^2-4\lambda_1\mu_1]}\},\\ w_2(s,x) &= (2\mu_1)^{-1}\{\alpha+\mu_1-\lambda_2 x+\sqrt{[(\alpha+\mu_1-\lambda_2 x)^2-4\lambda_1\mu_1]}\}.\end{aligned}\right\} \tag{2·16}$$

Since there is only one boundary condition, (2·15), the constant B may be set identically zero. Then the value of A appropriate to (2·15) is

$$A(s,y,x) = \frac{\mu_2(1-y^K x^{-1}) \sum_{n=1}^{\infty} x^n g_{01n} - y^K\{(1-y^{-1})\,\mu_2 g_{00}+1\}}{\mu_1(1-w_2)-\mu_2(1-y^{-1})}.$$

The unknown sum $\sum_{n=1}^{\infty} x^n g_{01n}$, a function of s and x only, can be found by substituting

$$y = \{1+(w_2-1)\,\mu_1\mu_2^{-1}\}^{-1}$$

in the numerator of this expression and equating it to zero;

$$\sum_{n=1}^{\infty} x^n g_{01n} = x\mu_2^{-1}\{\mu_1(w_2-1)\, g_{00}-1\}\{1-x[1+(w_2-1)\,\mu_1\mu_2^{-1}]^K\}^{-1}.$$

Writing

$$\beta(s,x) = \{w_2(s,x)-1\}\,\mu_1\mu_2^{-1}$$

the required generating function is

$$F(s;z,y,x) = \frac{(1-y^K)\, g_{00}}{1-zv_1} + \frac{xy\{y^K(1+\beta)^K-1\} + g_{00}\{\mu_1 y(x-y^K)(w_2-1) - \mu_2 y^K(1-y)\,[x(1+\beta)^K-1]\}}{(1-zw_1)\,[x(1+\beta)^K-1]\,[\mu_1 y(w_2-1)-\mu_2(1-y)]}. \tag{2·17}$$

The joint generating function of the numbers of priority and non-priority customers, irrespective of the phase of the current customer, is

$$F(s;z,1,x)=\frac{\mu_1(1-x)(w_2-1)g_{00}+x\{1-(1+\beta)^K\}}{\mu_1(1-zw_1)(w_2-1)\{1-x(1+\beta)^K\}}. \tag{2·18}$$

As a check, we note that the marginal distribution of the number of priority customers is the classical result

$$F(s;z,1,1)=\frac{1-w_1(s,1)}{s[1-zw_1(s,1)]}. \tag{2·19}$$

The generating function of particular interest, that of the non-priority customers, is

$$F(s;1,1,x)=\frac{\mu_1(1-x)(w_2-1)g_{00}+x\{1-(1+\beta)^K\}}{(s+\lambda_2-\lambda_2 x)\{1-x(1+\beta)^K\}}. \tag{2·20}$$

The solution is now complete except for the unknown $g_{00}(s)$, the Laplace transform of the null probability. An explicit expression for $g_{00}(s)$ may be found by the following well known argument. Since $F(s;1,1,x)$ is a generating function it must converge for at least $|x|\leqslant 1$, so that zeros in x of denominator and numerator within the unit circle coincide. By Rouché's Theorem the only zero of the denominator within the unit circle is that of

$$1-x\{1+\beta(s,x)\}^K=0.$$

If this zero is $x=\xi(s)$, then
$$g_{00}(s)=\{\mu_1(w_2(s,\xi)-1)\}^{-1}. \tag{2·21}$$

Using the Lagrange Inversion formula (Whittaker & Watson, 1950, p. 132) $\xi(s)$, and hence $g_{00}(s)$, can be found. The final result is too complicated for practical purposes and is not given here.

A case of especial interest is that when the service time is constant, that is $K=\infty$. Write $\mu_2=\nu_2 K$, so that the constant service time is of duration $1/\nu_2$. Then

$$\lim_{K\to\infty}\{1+\beta(s,x)\}^K=\exp\{\mu_1\nu_2^{-1}[w_2(s,x)-1]\}. \tag{2·22}$$

With this substitution the preceding formulae hold in this case also.

The equilibrium distributions, when they exist, are easily derived from (2·17). We use a generalization of Abel's Theorem (Widder, 1946, Chapter V) which yields

$$\lim_{s\to+0} sg_{rmn}(s)=\lim_{t\to\infty}P_{rmn}(t)=p_{rmn}$$

whenever the limit on the right-hand side exists. It is easy to see that the equilibrium distribution exists when $1>\rho_1+\rho_2=\lambda_1\mu_1^{-1}+\lambda_2 K\mu_2^{-1}$, and when this is so, we have directly from (2·18), the steady-state generating function

$$\lim_{s\to+0} sF(s;z,1,x)=\Phi(z,x)=\frac{\mu_1(1-x)p_{00}}{\{1-zw_1(0,x)\}\{1-x[1+\beta(0,x)]^K\}}. \tag{2·23}$$

Since $\Phi(1,1)=1$,
$$p_{00}=1-\rho_1-\rho_2. \tag{2·24}$$

Moments are easily found by differentiating $\Phi(z,x)$. The expectation of the number of non-priority customers, denoted by $\bar{n}_K$, is

$$\bar{n}_K=\frac{\rho_2\{2[1-\rho_1+\rho_1\mu_2(\mu_1 K)^{-1}]-(1-K^{-1})\rho_2\}}{2(1-\rho_1)(1-\rho_1-\rho_2)}. \tag{2·25}$$

In particular
$$\bar{n}_\infty=\frac{\rho_2\{2[1-\rho_1+\rho_1\nu_2\mu_1^{-1}]-\rho_2\}}{2(1-\rho_1)(1-\rho_1-\rho_2)}, \tag{2·26}$$

3. Model for breakdowns

The preceding model is now modified to describe in a more realistic manner the way in which breakdowns occur. The essential difference between a breakdown and a priority arrival is that breakdowns can occur only when a customer is in service, and also a queue of breakdowns cannot form. If r denotes the number of breakdowns, then r can take only the values 0 and 1. The random walk in the (r, n)-plane takes place only on the lines $r = 0$ and $r = 1$, excluding the point $(1, 0)$. If $r = 0$, the three permissible transitions from the point $(0, n)$, $n = 1, 2, \ldots$, are to the points $(0, n-1)$, $(0, n+1)$, or $(1, n)$. If $r = 1$ only two transitions are allowable, to $(1, n+1)$ or $(0, n)$. From the point $(0, 0)$ the only transition possible is to $(0, 1)$.

If $q_{rmn}(s)$ denotes the Laplace transform of the unknown probability $P_{rmn}(t)$ and $\alpha = s + \lambda_1 + \lambda_2$, then, for an empty system initially, the difference equations satisfied by $q_{rmn}(s)$ are

$$-(s+\lambda_2)\,q_{00} + \mu_2 q_{011} = -1 \quad (r = 0,\ n = 0), \tag{3·1}$$

$$-(\alpha+\mu_2)\,q_{0m1} + \mu_1 q_{1m1} + \mu_2 q_{0m+11} = 0 \quad (r = 0,\ n = 1,\ 1 \leqslant m < K), \tag{3·2}$$

$$-(\alpha+\mu_2)\,q_{0K1} + \lambda_2 q_{00} + \mu_1 q_{1K1} + \mu_2 q_{012} = 0 \quad (r = 0, n = 1, m = K), \tag{3·3}$$

$$-(\alpha+\mu_2)\,q_{0mn} + \lambda_2 q_{0mn-1} + \mu_1 q_{1mn} + \mu_2 q_{0m+1n} = 0 \quad (r = 0,\ n > 1,\ 1 \leqslant m < K), \tag{3·4}$$

$$-(\alpha+\mu_2)\,q_{0Kn} + \lambda_2 q_{0Kn-1} + \mu_1 q_{1Kn} + \mu_2 q_{01n+1} = 0 \quad (r = 0,\ n > 1,\ m = K), \tag{3·5}$$

$$-(s+\lambda_2+\mu_1)\,q_{1m1} + \lambda_1 q_{0m1} = 0 \quad (r = 1,\ n = 1,\ 1 \leqslant m < K), \tag{3·6}$$

$$-(s+\lambda_2+\mu_1)\,q_{1mn} + \lambda_1 q_{0mn} + \lambda_2 q_{1mn-1} = 0 \quad (r = 1, n > 1,\ 1 \leqslant m \leqslant K). \tag{3·7}$$

Proceeding as before the joint generating function of breakdowns and customers is

$$\begin{aligned} G(s; z, x) &= q_{00} + \sum_{r=0}^{1} \sum_{m=1}^{K} \sum_{n=1}^{\infty} z^r x^n q_{rmn}(s) \\ &= q_{00} - \frac{x(s+\mu_1+\lambda_2-\lambda_2 x+\lambda_1 z)\,[1-(1+\theta)^K]\,[(s+\lambda_2-\lambda_2 x)\,q_{00}-1]}{\mu_2(s+\mu_1+\lambda_2-\lambda_2 x)\,\theta[1-x(1+\theta)^K]}, \end{aligned} \tag{3·8}$$

where

$$\theta = \theta(s, x) = \frac{(s+\mu_1+\lambda_1+\lambda_2-\lambda_2 x)\,(s+\lambda_2-\lambda_2 x)}{\mu_2(s+\mu_1+\lambda_2-\lambda_2 x)}.$$

The generating function of the number of customers is

$$G(s; 1, x) = \frac{(1-x)\,(s+\lambda_2-\lambda_2 x)\,q_{00} - x[1-(1+\theta)^K]}{(s+\lambda_2-\lambda_2 x)\,[1-x(1+\theta)^K]}. \tag{3·9}$$

The null probability $\mathscr{L}^{-1}[q_{00}(s)]$ can be found using the Lagrange Inversion Theorem, but as in the preemptive priority model the calculations involved are lengthy.

If $\phi(x)$ is the generating function of the equilibrium distribution, then from (3·9),

$$\phi(x) = \frac{(1-x) f_{00}}{1-x[1+\theta(0, x)]^K}, \tag{3·10}$$

where f_{00} denotes the null probability. This is easily found to be

$$f_{00} = 1 - \rho_2(1+\rho_1), \tag{3·11}$$

which agrees with the result of White & Christie (1958, p. 90) in their study of the breakdown model for the negative exponential queue. Equation (3·11) implies that the condition for

the existence of the equilibrium distribution is $1 > \rho_2(1+\rho_1)$, which is less restrictive than the condition $1 > \rho_1+\rho_2$ for the preemptive model. The expected number of customers in the system, denoted by $\bar{b}_K$, is

$$\bar{b}_K = \frac{\rho_2\{2(1+\rho_1+\rho_1\rho_2\mu_2(\mu_1 K)^{-1})-(1-K^{-1})\rho_2(1+\rho_1)^2\}}{2(1-\rho_2-\rho_1\rho_2)}. \tag{3·12}$$

Formulae for the deterministic case, $K = \infty$, are given, as before, by rewriting μ_2 as $\nu_2 K$ and proceeding to the limit.

For purposes of comparison we list here some formulae for the Poisson input, Erlang service time queueing model in which no interruptions occur, either in the form of breakdowns or priority arrivals. If λ_2, μ_2 are respectively the arrival and service parameters; $h_0(s)$ the Laplace transform of the null probability; and if initially there are no customers in the system, then the generating function is

$$H(s,x) = \frac{(1-x)(s+\lambda_2-\lambda_2 x)h_0+x(1-\gamma^K)}{(s+\lambda_2-\lambda_2 x)(1-x\gamma^K)}, \tag{3·13}$$

where

$$\gamma(s,x) = (s+\mu_2+\lambda_2-\lambda_2 x)\mu_2^{-1}.$$

The null probability $\mathscr{L}^{-1}[h_0(s)]$ can be found explicitly in this case in terms of generalized Bessel functions (Luchak, 1958). The equilibrium generating function is

$$\psi(x) = \frac{(1-x)(1-\rho_2)}{1-x\{1+(1-x)\lambda_2\mu_2^{-1}\}^K} \tag{3·14}$$

and the expected queue length

$$\bar{a}_K = \frac{\rho_2\{2-(1-K^{-1})\rho_2\}}{2(1-\rho_2)}. \tag{3·15}$$

If the rate at which interruptions occur, λ_1, is zero, then putting $\mu_1 = \mu_2$, it is easy to check that the formulae derived for the priority and breakdown models reduce to (3·13)–(3·15).

References

Barry, J. Y. (1956). A priority queueing problem. *Opns. Res.* **4**, 385.

Heathcote, C. R. (1959). The time-dependent problem for a queue with preemptive priorities. *Opns. Res.* **7**, 670–80.

Heathcote, C. R. (1960). A simple queue with several preemptive priority classes. To be published.

Luchak, G. (1958). The continuous time solution of the equations of the single channel queue with a general class of service-time distributions by the method of generating functions. *J. R. Statist. Soc.* B, **20**, 176–81.

Morse, P. M. (1958). *Queues, Inventories and Maintenance.* New York: John Wiley and Sons.

Stephan, F. (1958). Two queues under preemptive priority. *Opns. Res.* **6**, 399–418.

Whittaker, E. T. & Watson, G. N. (1950). *Modern Analysis*, 4th edition. Cambridge.

White, H. & Christie, L. S. (1958). Queueing with preemptive priorities or with breakdown. *Opns. Res.* **6**, 79–95.

Widder, D. V. (1946). *The Laplace Transform.* Princeton University Press.

Biometrika (1961), **48**, 1 *and* 2, *p.* 65
Printed in Great Britain

A two-sample sequential *t*-test

BY J. HAJNAL
London School of Economics

1. THE BASIS OF THE TEST

Observations are taken from two normal populations with (unknown) means μ_x and μ_y and common (unknown) variance σ^2. It is desired to test the hypothesis (H_0) that

$$\mu_x = \mu_y$$

against the alternative hypothesis (H_1) that

$$\frac{(\mu_x - \mu_y)^2}{\sigma^2} = \zeta^2.$$

The ordinary non-sequential test in this situation is, of course, the two-sample t-test. Let the observations from the two populations be denoted respectively x_j and y_j ($j = 1, 2, \ldots$). Suppose that N_x observations have been taken from one population and N_y from the other. We compute

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{u}}, \tag{1}$$

where

$$\bar{x} = \frac{1}{N_x}\Sigma x_j, \quad \bar{y} = \frac{1}{N_y}\Sigma y_j,$$

$$s^2 = \frac{\sum_j (x_j - \bar{x})^2 + \sum_j (y_j - \bar{y})^2}{N_x + N_y - 2}$$

and

$$u = \frac{1}{N_x} + \frac{1}{N_y}.$$

Under H_0, t as computed according to (1) has the t-distribution with $N_x + N_y - 2$ degrees of freedom. Under H_1 on the other hand it has the non-central t-distribution with $N_x + N_y - 2$ degrees of freedom and non-centrality parameter $\Delta = \zeta/\sqrt{u}$. The non-central t-distribution (see, for example, Johnson & Welch, 1939) is the distribution of

$$\frac{\epsilon + \Delta}{\sqrt{W_f}},$$

where Δ is a constant, ϵ is a normal variable with mean 0 and variance 1, W_f is distributed as χ_f^2/f (χ_f^2 having f degrees of freedom).

In order to obtain a sequential t-test for this situation we compute t at each stage, using all the observations already accumulated, and calculate the likelihood ratio

$$\frac{p(t^2|H_1)}{p(t^2|H_0)}, \tag{2}$$

where $p(t^2|H_i)$ is the probability density of t^2 under the hypothesis H_i. Sampling proceeds as long as the inequality

$$\frac{\beta}{1-\alpha} < \frac{p(t^2|H_1)}{p(t^2|H_0)} < \frac{1-\beta}{\alpha} \tag{3}$$

holds. (α and β are the risks of falsely accepting H_0 and H_1, respectively.) When (3) is violated, no further observations are taken. Whether H_0 or H_1 is accepted, depends on which side of the inequality is violated.

The ratio (2) is a ratio of the density function of non-central t^2 to that of central t^2. It is known (e.g. Introduction to National Bureau of Standards, 1951) that such a ratio can be expressed in the form

$$\exp(-\tfrac{1}{2}\Delta^2)\, F\left[\frac{f+1}{2}, \frac{1}{2}; \frac{\Delta^2 t^2}{2(f+t^2)}\right], \tag{4}$$

where Δ is the non-centrality parameter, f is the number of degrees of freedom and $F(a, b; x)$ is the confluent hypergeometric function.

Using expression (4), the inequality (3) becomes

$$\frac{\beta}{1-\alpha} < e^{-\zeta^2/2u} F\left[\frac{N-1}{2}, \frac{1}{2}; \frac{\zeta^2 t^2}{2u(N-2+t^2)}\right] < \frac{1-\beta}{\alpha}, \tag{5}$$

where $N = N_x + N_y$. It is usually more convenient to take logarithms and write

$$\ln\frac{\beta}{1-\alpha} < -\frac{\zeta^2}{2u} + \ln F\left[\frac{N-1}{2}, \frac{1}{2}; \frac{\zeta^2 t^2}{2u(N-2+t^2)}\right] < \ln\frac{1-\beta}{\alpha}. \tag{6}$$

We now need to show that the procedure described constitutes a valid sequential probability ratio test as developed by Wald (1947).

We suppose first that at each stage of sampling a predetermined number of x- and y-observations are taken. Thus we may take observations singly, alternating between the two populations, or one x- and one y-observation at each stage, or one x- and three y-observations, etc. (The effects of taking observations in groups are discussed in Wald's book.)

Denote the sequence of the first N values observed $x_1, x_2, \ldots, x_N$; $y_1, y_2, \ldots, y_N$ by S_N and let $P(S_N|H_j)$ be the joint probability density function for S_N under hypothesis H_j. $p(S_N|H_j)$ will involve the parameters μ_x, μ_y and ζ of which jointly sufficient estimators are available. These parameters may be expressed in terms of $|\zeta|$, σ, μ_x and the sign of ζ. The distribution of t^2 depends only upon $|\zeta|$, but not upon the other parameters.

By a theorem due to Cox (1952)†, we may write

$$p(S_N|H_j) = p(t^2|H_j) \times R \quad (j = 0, 1), \tag{7}$$

where R does not involve ζ.

Hence

$$\frac{p(S_N|H_1)}{p(S_N|H_0)} = \frac{p(t^2|H_1)}{p(t^2|H_0)}. \tag{8}$$

It may then be shown by Cox's argument that the test proposed has probabilities of error roughly equal to α and β, provided the probability is one that the procedure terminates.

The certainty of termination may be demonstrated by comparing our test with a suitably constructed single-sample two-sided sequential t-test. We assume that our sampling procedure is such that in each batch of N' successive observations there are rN' x-observations and $(1-r)N'$ y-observations and we consider the probability that the inequality (3) is violated when $N = N', 2N', 3N', \ldots$ observations have been analysed. (This probability is

† The main difficulty in checking the applicability of Cox's theorem arises from the last of his conditions. The set of transformations which carries an observation x_j into $a+bx_j$ and y_j into $a+by_j$ $(-\infty < a < \infty;\ -\infty < b < \infty)$ satisfies this condition.

smaller than the probability that the test terminates at some point up to and including the Nth observation.) We suppose that our two sample test is carried out in a situation where

$$\frac{(\mu_x-\mu_y)^2}{\sigma^2} = (\zeta')^2 \text{ say.}$$

Now consider a single-sample two-sided t-test of the hypothesis H_0 that the mean μ of a normal population of variance σ^2 is zero against H_1 that $\mu^2/\sigma^2 = \zeta^2 r(1-r)$. The risks of error have the same values α and β as before. Suppose the test is applied to observations drawn from a normal population with variance σ^2 and mean μ such that

$$\mu^2/\sigma^2 = (\zeta')^2 r(r-1).$$

At each stage t is computed. After N observations this t will have the non-central t-distribution with $N-1$ degrees of freedom and non-centrality parameter $\zeta'\sqrt{\{Nr(1-r)\}}$. From the observed t we calculate the likelihood ratio

$$\frac{p(t^2|H_1)}{p(t^2|H_0)}$$

using in the numerator a non-central t-distribution with $N-1$ degrees of freedom and non-centrality parameter $\zeta\sqrt{\{Nr(1-r)\}}$, while the density function for the denominator is the ordinary t-distribution with $N-1$ degrees of freedom. With this value of the likelihood ratio we then observe whether the inequality (3) is violated.

Compare these steps with the proposed two-sample t-test bearing in mind that, for values of N which are multiples of N',

$$1/u = Nr(1-r).$$

The t computed with N observations has the same distribution as the t in the single-sample test except that there is 1 degree of freedom less. The likelihood ratio is also based on the same distributions except that in the two-sample test there are only $N-2$ degrees of freedom. If the single-sample test ends with probability one, it will then follow that the same is true of the two-sample test.

The certainty of termination for the two-sided single-sample t-test seems to be generally accepted. It may be proved by considering two one-sided single sample t-tests. For a two-sided test of the hypothesis H_0 that $\mu = 0$, against H_1 that $\mu^2/\sigma^2 = \delta^2$ with α and β as the risks of falsely accepting H_0 and H_1, respectively, we take the following one-sided tests, both of which are assumed to be carried out on the same set of observations as the corresponding two-sided test

$$H_0\colon \mu = 0 \quad \text{against} \quad H_1\colon \mu/\sigma = -\delta,$$
$$H_0\colon \mu = 0 \quad \text{and against} \quad H_1\colon \mu/\sigma = +\delta.$$

For both these one-sided tests the risk of falsely accepting H_0 is set at α as before, while the risk of falsely accepting H_1 is taken as $\frac{1}{2}\beta$. Then it may be shown that the two-sided test will terminate no later than the point at which both the one-sided tests have ended. The fact that both these tests will end with probability one follows from the paper by David & Kruskal (1956).

To sum up we may conclude that the proposed two-sample test will end with probability one and that it is a valid sequential probability ratio test.

We now abandon the assumption that at each stage a predetermined number of observations are taken from each population. Suppose, for example, that a medical researcher is concerned with the difference between men and women patients in a certain characteristic. He might then wish to take each successive patient, male or female, who becomes available and decide at each stage whether additional observations are needed. To cover such a situation, suppose there is a some constant probability π (known or unknown) that an observation be taken from the x-population.

The probability frequency function of the observations may be written

$$\pi^{N_x}(1-\pi)^{N_y}\, p(S_N|H_j),$$

where $p(S_N|H_j)$ is now the conditional probability of the particular values $(x_1, x_2, \ldots, x_{N_x}; y_1, y_2, \ldots, y_N)$ given the sequence in which observations have been drawn from either the x- and or the y-population. The factor

$$\pi^{N_x}(1-\pi)^{N_y}$$

is independent of H_j and cancels out in the likelihood ratio. To show that the test is certain to end, we note that, with probability one, as $N \to \infty$,

$$1/u \to N\pi(1-\pi).$$

By using this fact in combination with the method of proof sketched earlier for the case where the sequence of x- and y-observations is predetermined, it may be shown that the test will end with probability one. (In carrying out this proof we take $\pi(1-\pi)$ in place of $r(1-r)$).

The proposed test may, by an extension of these arguments, be shown to be valid even for cases where the probability that the next observation be drawn from the x- or y-population changes as the sampling proceeds. Let π_N be the probability that the Nth observation is from the x-population. Then the proposed test is valid if π_N is the same under H_0 and H_1, is independent of all earlier observations and there exist numbers $0 < \pi_{\min.}$ and $\pi_{\max.} < 1$ such that for all N

$$\pi_{\min.} < \pi_N < \pi_{\max.}.$$

The theory which has been outlined is an extension of that developed by Rushton (1952) and Colombo (1959). In Rushton's paper the crucial formulae relating to the two-sample case contain misprints.† The techniques for practical application and estimation of average sample size which are described in 2–3 below are new.

2. PRACTICAL COMPUTATION

Running totals of Σx_i, Σy_i and $(\Sigma x_i^2 + \Sigma y_i^2)$ can be accumulated as the observations come in. If we write

$$T^2 = \frac{(\bar{x}-\bar{y})^2}{\Sigma x_i^2 + \Sigma y_i^2 - \left(\dfrac{(\Sigma x_i)^2}{N_x} + \dfrac{(\Sigma y_i)^2}{N_y}\right)} = \frac{(\bar{x}-\bar{y})^2}{S^2} \tag{9}$$

it follows that

$$\frac{t^2}{f+t^2} = \frac{T^2}{u+T^2}.$$

T^2 and hence $T^2/(u+T^2)$ required in (6) may be computed instead of expressions involving t. Once this is done there are various ways of checking whether equation (6) still holds. The logarithm of the hypergeometric function can be evaluated by Rushton's approximation

† The corrigenda may be found on an unnumbered page wrongly placed before the first page (p. 287) of parts 3–4 of *Biometrika*, vol. 41 (1954). See vol. 42 (1955), p. 277.

(Rushton, 1952) or by the tables of the confluent hypergeometric function published by the U.S. National Bureau of Standards (1949).

The simplest procedure, however, is to utilize another set of tables prepared by the National Bureau of Standards (1951). These tables, which will be referred to as the 'N.B.S. tables', are for use with the single sample sequential t-test. For selected L, n and δ they give values of Z which are solutions of the equation

$$L = \ln F[\tfrac{1}{2}n, \tfrac{1}{2}; \tfrac{1}{2}Z\delta^2] - \tfrac{1}{2}n\delta^2. \tag{10}$$

The procedure for the two sample t-test may therefore be put as follows. At each stage calculate the T^2 of (9) and compute

$$Z_N = \frac{(N-1)\,T^2}{u+T^2}. \tag{11}$$

Continue sampling so long as $$\underline{Z}_N < Z_N < \bar{Z}_N. \tag{12}$$

The limits $\underline{Z}_N$ and $\bar{Z}_N$, may be obtained from the N.B.S. tables, entered with

$$n = N-1,$$

$$\delta = \frac{\zeta}{\sqrt{\{u(N-1)\}}} \tag{13}$$

and $$L = \ln\frac{\beta}{1-\alpha} \ (\text{for } \underline{Z}_N), \quad L = \ln\frac{1-\beta}{\alpha} \ (\text{for } \bar{Z}_N). \tag{14}$$

(For convenience we shall write ζ in place of $|\zeta|$ from now on.)

Table 1. *Observations and preliminary calculations*

	x-observations					*y*-observations				
N	N_x	x_i	Σx_i	$\bar{x}$	$\frac{(\Sigma x_i)^2}{N_x}$	N_y	y_i	Σy_i	$\bar{y}$	$\frac{(\Sigma y_i)^2}{N_y}$
1	1	− 4·0	− 4·0	—	—		—	—	—	—
2	—	—	—	—	—	1	+ 6·4	+ 6·4	—	—
3	—	—	—	—	—	2	+23·0	+29·4	—	—
4	2	− 3·8	− 7·8	—	—		—	—	—	—
5	—	—	—	—	—	3	− 4·9	+24·5	+ 8·17	200·08
6	3	+ 4·4	− 3·4	−1·13	3·85		—	—	—	—
7	4	+ 5·3	+ 1·9	+0·475	0·90		—	—	—	—
8	—	—	—	—	—	4	+23·8	+48·3	+12·075	583·22
9	5	+10·0	+11·9	+2·38	28·32		—	—	—	—
10	—	—	—	—	—	5	+16·3	+64·6	+12·92	834·63
11	6	+15·7	+27·6	+4·10	126·96		—	—	—	—
12	7	+30·1	+57·7	+8·24	475·61		—	—	—	—

A way of setting out the calculations is illustrated in Tables 1–3. For the example shown observations were drawn at random from a normal population with mean and variance both equal to 10. The observations were randomly allocated to the two groups with equal probability. This corresponds to a situation where observations come in with roughly equal frequency, but not in strict alternation, from two populations whose means are in fact equal, i.e. $\pi = 0{\cdot}5$ and H_0 is true. The alternative hypothesis tested is that $\zeta = 2$, with $\alpha = \beta = 0{\cdot}05$.

As observations become available they are entered on the appropriate side of Table 1, where some additional calculations are also performed. The computation of stopping limits

is shown in Table 2. Finally, in Table 3, figures from the two sides of Table 1 are combined to calculate the sequential criterion Z_N which can be compared with the stopping limits obtained in Table 2. In each table a new line is added for each observation. The procedure is governed by the N.B.S. tables and can be understood only by reference to them in conjunc-

Table 2. *The calculation of stopping limits*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Lower limit			Upper limit		
			Nearest N.B.S. limits		Interpolated limit	Nearest N.B.S. limits		Interpolated limit
N	$u=\frac{1}{N_x}+\frac{1}{N_y}$	$\frac{2\cdot0}{\sqrt{[u(N-1)]}}=\delta$	$\delta=1\cdot0$	$\delta=1\cdot2$	$\underline{Z}_N$	$\delta=1\cdot0$	$\delta=1\cdot2$	$\overline{Z}_N$
2	2·0	1·41	—	—	—	—	—	—
3	1·5	1·15	—	—	—	—	—	—
4	1·0	1·15	—	—	—	—	—	—
5	0·833	1·10	—	—	—	—	—	—
6	0·666	1·10	—	(0·22)	—	4·98	4·48	4·73
7	0·583	1·07	0·02	0·45	0·17	4·99	4·63	4·86
8	0·50	1·07	0·19	0·68	0·36	5·05	4·81	4·97
9	0·45	1·05	0·35	0·92	0·49	5·14	5·01	5·11
10	0·40	1·05	0·52	1·17	0·68	5·26	5·22	5·25
11	0·367	1·04	0·69	1·42	0·84	5·39	5·45	5·40
12	0·343	1·03	0·87	1·67	0·99	5·53	5·68	5·55

Notes:

Column (3). Calculation of δ by formula (13) for entry in N.B.S. tables.

Columns (4)–(5). Limits from N.B.S. tables under $L=-\ln 19$ since $\alpha=\beta=0\cdot05$.

Column (6). Linear interpolation between columns (4) and (5) for δ computed under (3), e.g. $0\cdot02+0\cdot07/0\cdot20\ (0\cdot45-0\cdot02)=0\cdot17$.

Columns (7)–(8). Limits from N.B.S. tables under $L=\ln 19$.

Column (9). Linear interpolation between columns (7) and (8).

Table 3. *Computation of sequential criterion*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	$\Sigma(x_i^2+y_i^2)$	S^2	$\bar{y}-\bar{x}$	T^2	$u+T^2$	Z_N	$\underline{Z}_N$	$\overline{Z}_N$
1	16·00	—	—	—	—	—	—	—
2	56·96	—	—	—	—	—	—	—
3	585·96	—	—	—	—	—	—	—
4	600·40	—	—	—	—	—	—	—
5	624·41	—	—	—	—	—	—	—
6	643·77	439·83	+ 9·30	0·197	0·863	1·14	—	4·73
7	671·86	470·87	+ 7·69	0·126	0·709	1·07	0·17	4·86
8	1 238·30	656·18	+11·60	0·205	0·705	2·04	0·36	4·97
9	1 338·30	726·76	+ 9·70	0·129	0·579	1·78	0·49	5·11
10	1 603·99	741·04	+10·54	0·150	0·550	2·45	0·68	5·25
11	1 850·48	888·89	+ 8·32	0·0779	0·445	1·75	0·84	5·40
12	2 756·49	1 446·25	+ 4·68	0·0151	0·359	0·46	0·99	5·55

Notes:

Column (3). $$S^2=\Sigma(x_i^2+y_i^2)-\frac{(\Sigma x_i)^2}{N_x}-\frac{(\Sigma y_i)^2}{N_y}.$$

Column (5). $$T^2=\frac{(\bar{y}-\bar{x})^2}{S^2}.$$

Column (7) $$Z_N=\frac{(N-1)\,T^2}{u+T^2}.$$

tion with equations (11) to (14) above. In the present example at $N = 12$ we find that $Z_N < \underline{Z}_N$. Hence sampling stops with the (correct) decision to accept H_0.

Column (2) shows the computation of the value of δ to be used for entry into the N.B.S. tables in accordance with formula (13). Normally the δ so calculated will vary from line to line and will not correspond to any of the δ's provided for in the N.B.S. tables. To find fairly precise stopping limits interpolation is, therefore, necessary. Linear interpolation is almost invariably sufficient. (The accuracy of interpolation with respect to δ is discussed on pages xii and xvii of the introduction to the N.B.S. tables.) Some of the interpolated limits in Table 2 may be in error in the second place of decimals, but this is of no practical significance.

$\underline{Z}_N$ and $\overline{Z}_N$ are only given in the N.B.S. tables from certain values of N onwards; i.e. whatever the observations turn out to be, it is necessary for N to exceed a certain minimum before it is possible to obtain a likelihood ratio which violates the inequality (3) and permits sampling to stop. Mathematically the reason is that for low N one side of the inequality can be violated only if Z_N is negative and the other if Z_N exceeds $N-1$; neither of these alternatives is possible. Sometimes, as in the example shown, a difficulty arises in determining one of the initial stopping points in Table 2. The N.S.B. tables give a value of $\underline{Z}_N$ for $N = 6$, $\delta = 1{\cdot}2$, but under $\delta = 1{\cdot}0$ the first $\underline{Z}_N$ corresponds to $N = 7$. Is there a $\underline{Z}_N$ for $N = 6$ and $\delta = 1{\cdot}1$? It is not difficult to compute an appropriate $\underline{Z}_N$, but the simplest solution is to adopt the rule that acceptance of H_0 first becomes possible at the N for which $\underline{Z}_N$ is given in the N.B.S. tables for both neighbouring values of δ. A corresponding rule is, of course, appropriate for determining the smallest sample size for the acceptance of H_1. The consequences of these rules will be a negligible increase in the average number of observations needed; the risk of either kind of error will not exceed the levels specified.

The interpolation of the precise stopping limits between those given in the N.B.S. tables is usually only necessary for one or two observations at the end. In the example shown, no interpolation at all would have been required; a comparison of Z_N with the limits provided by the N.B.S tables would have been adequate to establish at every stage whether sampling should continue. Much of Tables 1 and 3 need be computed only when N is large enough so that it is possible to stop sampling.

If the sequence in which observations will be drawn from the two groups is fixed in advance, Table 2 can be computed before the experiment and the same interpolated stopping limits can then be used on several occasions. A simpler procedure which may often be practical is to use the N.B.S. limits as they stand for a value of δ which provides a test having approximately the desired characteristics. For example, in a trial where observations are drawn in strict alternation from each of the two populations and it is desired to have roughly $\zeta = 2$ as in the example illustrated, one might well use the N.B.S. limits for $\delta = 1$. As may be inferred from column (3) of Table 2 this would provide a slightly more stringent test than intended, i.e. it would have the characteristics of a test in which ζ or the risks of error were slightly smaller than those selected. However, since both ζ and the risks of error have often to be selected in somewhat arbitrary fashion, this may be no inconvenience.

In applications where observations from the two samples come in with roughly equal frequency, it is possible to take the observations in pairs and then use $x_1 - y_1, x_2 - y_2$, etc., as successive values for a single sample sequential t-test (putting $\delta = \zeta/\sqrt{2}$.) More generally, if the observations are taken in batches of N' at a time of which $\pi N'$ are from the x-population and $(1-\pi)N'$ from the y-population, we may compute $(1-\pi)\Sigma x_i - \pi\Sigma y_i$ at each stage. (Σx_i and Σy_i here stand for the sums of all the x and of all the y observations in one batch.)

A single sample t-test with these values using $\delta = \zeta\sqrt{\{N'\pi(1-\pi)\}}$ is then equivalent to the two-sample procedure.

The single sample test is simpler to carry out and the advantage of the full two-sample procedure have to be weighed against the additional labour. The possibility of earlier stopping in the full procedure arises mainly because the degrees of freedom for estimating σ^2 accumulate faster, but also because the process is examined more often and because, where the selection of observations is not under the experimenter's control, there will usually not be exactly $\pi N'$ and $(1-\pi)N'$ observations of the two kinds in successive batches of N' observations. Where the collection of observations costs much time and effort, as for example in clinical medicine, the additional work of the full two-sample analysis will often be justified, especially where it is important to be able to stop early if the difference between the means compared appears large. When two treatments are compared in a clinical trial this is very important, since to continue with a trial when large differences have been revealed would imply exposing some patients to the risk of inferior treatment.

The advantage of the full two-sample procedure in this respect can be substantial. For example, for $\zeta = 2, \pi = 0{\cdot}2$ the full two-sample procedure may, if the observations so fall out, end with a decision that there is a difference between the means when $N = 6$. If 5 observations are taken as a batch each time, such a decision cannot take place before 4 batches, i.e. 20 observations have been examined.

Situations where it is not possible to take observations in equal numbers from the two populations being compared, or where the experimenter cannot control the selection of observations from one population or the other arise in various ways in clinical medicine. For example, in a trial of a new drug of which only small quantities are available, it may be desired to compare a small number of patients treated with the drug with a larger number of controls; one of the two groups being compared in a study may be patients suffering from a rare disease, etc. The present method was in fact devised in response to the needs of a clinical trial.

3. AVERAGE SAMPLE SIZE

There is no known method of computing the average number of observations needed for sequential tests of composite hypotheses, but it seems worth while to apply to our tests a conjecture due to Bhate (1955).

We consider only the case where H_0 is true. Let λ_N be the logarithm of the likelihood ratio computed on N observations, i.e.

$$\lambda_N = \ln \frac{p(S_N|H_1)}{p(S_N|H_0)}. \tag{15}$$

We suppose that repeated samples of fixed size (N observations) are taken and compute $E(\lambda_N)$. $E(\lambda_N)$ is, of course, a function of N. Under H_0, $E(\lambda_N)$ will decrease with increasing N. Bhate's conjecture says that the value of N, say $\overline{N}$, for which

$$E(\lambda_{\overline{N}}) = (1-\alpha)\ln\left(\frac{\beta}{1-\alpha}\right) + \alpha \ln\left(\frac{1-\beta}{\alpha}\right) \tag{16}$$

is an approximation to the average sample number of our test.

Bhate's conjecture is a natural extension to composite hypotheses of Wald's well-known formula giving the average sample number for a sequential probability ratio test of a simple hypothesis. Ray (1956) has collected a number of instances where Bhate's conjecture is in good agreement with the results of sampling experiments.

To apply Bhate's conjecture we have to take the expected value of

$$\lambda_N = -\frac{\zeta^2}{2u} + \ln F\left[\frac{N-1}{2}, \frac{1}{2}; \frac{\zeta^2 t^2}{2u(N-2+t^2)}\right] \tag{17}$$

for fixed N.

We consider the case† where the order in which observations are taken from each of the populations is not fixed, but there is a probability π that an observation comes from the x population. In (17) not only t^2, but also $u = 1/N_x + 1N_y$ is thus a random variable, though $N_x + N_y = N$ is assumed fixed. We have

$$E\left(\frac{1}{u}\right) = \pi(1-\pi)(N-1). \tag{18}$$

We follow Ray (1956) in approximating to the expected value of $\ln F[\frac{1}{2}(N-1), \frac{1}{2}; X]$ by $\ln F[\frac{1}{2}(N-1), \frac{1}{2}; E(X)]$. If t^2 has the F-distribution with 1 and f degrees of freedom, $t^2/(f+t^2)$ has a β-distribution, and

$$E\left(\frac{t^2}{N-2+t^2}\right) = \frac{1}{N-1}. \tag{19}$$

By virtue of the independence of u and t^2 our problem reduces to solving for $\overline{N}$ the equation

$$-\tfrac{1}{2}[\zeta^2\pi(1-\pi)(\overline{N}-1)] + \ln F[\tfrac{1}{2}(\overline{N}-1), \tfrac{1}{2}; \tfrac{1}{2}\zeta^2\pi(1-\pi)]$$
$$= (1-\alpha)\ln\left(\frac{\beta}{1-\alpha}\right) + \alpha\ln\left(\frac{1-\beta}{\alpha}\right). \tag{20}$$

Solutions of this equation for $\alpha = \beta = 0{\cdot}05$ and varying values of ζ and π are given in Table 4. The equation is easily solved by means of the tables of the confluent hypergeometric function published by the National Bureau of Standards (1949) provided that an approximate value of $\overline{N}$ is available for a starting point. The starting point was obtained by noting that in these tables the function tabulated, namely

$$\frac{1}{\sqrt{(2n\overline{X})}} \ln F(\tfrac{1}{2}n, \tfrac{1}{2}; X)$$

is never far from 0·8 over the range likely to be relevant. Putting this value into equation (20), we obtain a quadratic equation for $\overline{N}-1$ and hence, for $\alpha = \beta = 0{\cdot}05$, approximately

$$\overline{N} - 1 = \frac{10{\cdot}5}{\zeta^2\pi(1-\pi)}. \tag{21}$$

Table 4 was obtained by solving equation (20). A comparison of this table with (21) shows that this approximation is sufficiently good in all cases where $\overline{N}$ is not below about 10 to make it pointless to solve equation (20) precisely. When the mean sample size is small, Bhate points out that his conjecture, like Wald's original formula from which it is derived, will tend to be an underestimate.

The formula just stated should be compared with that which gives the number of observations required for the equivalent test with a sample of fixed size, on the assumption that the variance is known. Suppose we wish to test the difference between the means of two populations having a common variance σ^2. We used a two-sided significance test at the 5 % level.

† The case where the number of x and y observations for any N total observations is known in advance is, of course, simpler. It turns out that in the approximation (21) N takes the place of $(\overline{N}-1)$.

We take a sample of, say, N^* observations, $N^*\pi$ being from one population and $N^*(1-\pi)$ from the other, where π is a given fraction. How large must N^* be if the test is to have a 95 % chance of giving a 'significant' result, when the true difference between the means is $\zeta\sigma$? The answer is provided by the formula

$$N^* = \frac{13}{\zeta^2\pi(1-\pi)}. \tag{22}$$

The comparison between the two formulae (i.e. the fraction 10·5/13) is not a full measure of the saving in the number of observations provided by the sequential analysis, for two reasons. First formula (22) assumes that the variance is known. For a t-test of equivalent strength where $\pi\zeta^2(1-\pi)$ is large, a somewhat greater N^* would be required (e.g. for $\pi\zeta^2(1-\pi) = 1$ formula (22) gives $N^* = 13$, using the non-central t-distribution we obtain $N^* = 15$). Secondly, in a situation where observations cannot be freely taken from either population, it will in most cases take more than N^* observations before πN^* x-observations and $(1-\pi)N^*$ y-observations have been assembled.

Table 4. *Average sample size under* H_0

(Bhate's conjecture; $\alpha = \beta = 0{\cdot}05$.)

ζ \ π	0·1	0·2	0·5
1	117	66	43
2	30	18	12
3	14	9	6
4	9	6	4

For values of α and β other than 0·05 the constants in the numerators of (21) and (22) would, of course, be different from those given. For $\alpha = \beta = 0{\cdot}01$, the numerator of formula (21) would be 16 and that of formula (22) would be 24.

Finally, it seems appropriate to record the application of Bhate's conjecture to the single sample sequential t-test. The average sample size, $\overline{N}$, if the true mean is zero, is the solution of the equation

$$-\tfrac{1}{2}\overline{N}\delta^2 + \ln F[\tfrac{1}{2}\overline{N}, \tfrac{1}{2}; \tfrac{1}{2}\delta^2] = (1-\alpha)\ln\left(\frac{\beta}{1-\alpha}\right) + \alpha\ln\left(\frac{1-\beta}{\alpha}\right). \tag{23}$$

By comparison with formula (20) it is clear that a good approximation to the solution, for $\alpha = \beta = 0{\cdot}05$, is provided by $10{\cdot}5/\delta^2$.

I am grateful to Dr N. L. Johnson for drawing my attention to Bhate's conjecture and making very helpful suggestions for the revision of the manuscript. I should also like to thank the referee for pointing out an important gap in the argument.

REFERENCES

BHATE, D. H. (1955). Sequential analysis with special reference to distributions of sample size. London University Ph.D. thesis (unpublished).

COLOMBO, B. (1959). Appunti di metodologia sequenziale. *Memorie della Accademia Patavina di SS. LL. AA. Classe di Scienze Matematiche e Naturali*, **71**, 3–30.

COX, D. R. (1952). Sequential tests for composite hypotheses. *Proc. Camb. Phil. Soc.* **48**, 290–9.

DAVID, H. T. & KRUSKAL, W. H. (1956). The WAGR sequential t-test reaches a decision with probability one. *Ann. Math. Statist.* **27**, 797–805.

JOHNSON, N. L. & WELCH, B. L. (1939). Applications of the non-central *t*-distribution. *Biometrika*, **31**, 362–89.

NATIONAL BUREAU OF STANDARDS (1949). *Tables of the Confluent Hypergeometric Function* $F(\frac{1}{2}n, \frac{1}{2}; x)$ *and Related Functions. Applied Mathematics Series*, no. 3. Washington: Government Printing Office.

NATIONAL BUREAU OF STANDARDS (1951). *Tables to Facilitate Sequential t-tests. Applied Mathematics Series*, no. 7. Washington: Government Printing Office.

RAY, W. D. (1956). Sequential analysis applied to certain experimental designs in the analysis of variance. *Biometrika*, **43**, 388–403.

RUSHTON, S. (1952). On a two-sided sequential *t*-test. *Biometrika*, **39**, 302–8.

WALD, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons, Inc.

Biometrika (1961), **48**, 1 *and* 2, *p.* 77
Printed in Great Britain

Absolute and incomplete moments of the multivariate normal distribution

BY S. NABEYA
Hitotsubashi University, Tokyo

1. *Introduction.* In general it is difficult to evaluate the exact sampling distribution of a statistic which is a function of absolute values of some correlated variates, even when the parent distribution is normal. In such cases it is usual to evaluate the moments of the statistic, and to approximate the sampling distribution by, e.g. a Pearson type distribution, a Gram–Charlier series, or the distribution of some power of a variable which has a gamma distribution, etc. In these cases it is often necessary to evaluate the absolute moments of the multivariate normal distribution. Geary (1936) was the first to have worked according to this idea.

The present author evaluated some of the absolute moments of the normal distribution with means zero, and in (1951) gave the results for the bivariate case up to 12th order. In (1952) he gave an integral formula for evaluating absolute moments from the characteristic function in the general multivariate distribution, and using this formula he derived the absolute moments of the trivariate normal distribution with means zero, also up to 12th order.

Working independently, Kamat (1953*a*) derived the same absolute moments for the bivariate and trivariate cases, and further obtained the results for the four-variate case by expansion in series. His method was to evaluate contributions to the absolute moment from each orthant, named incomplete moments, and sum up all these contributions. Incomplete moments were used also for other purposes. He argued that my method was not applicable to the evaluation of incomplete moments, and that his method was simpler than mine in evaluating certain absolute moments. Using the formulae for absolute moments, he gave the third moment of the mean deviation (1954), and of Gini's mean difference (1953*c*), in samples from a normal population, and further he contributed to the sampling distribution of some statistics involving absolute values.

To approximate the sampling distribution, it is often necessary to evaluate the moments up to the fourth order. Thus for a statistic which is a linear form of absolute values of normally correlated variables as in Geary's (1936) case or in Kamat's (1953*b*) case etc., Nabeya's (1951, 1952) and Kamat's (1953*a*) results given in exact form are not sufficient. It is desirable to evaluate $E[|x_1x_2x_3x_4|]$ in exact form, where x_1, x_2, x_3, x_4 are distributed according to the four-variate normal distribution with means zero. In Geary's (1936) and Kamat's (1953*a*) papers, the absolute moment of this form was found as a series expansion. Kamat (1953*b*) calculated $E[|x_1x_2x_3x_4|]$ in a special case of the correlation matrix, on the one hand in exact form using my integral formula, on the other hand in approximate form using his series expansion. He pointed out that the evaluation by my method was very elaborate even in his special case.

In § 2 of this paper I shall give an extension of the integral formula given in Nabeya (1952) for the evaluation of absolute moments, and I shall note that the extended formula may be used for evaluating incomplete moments. In § 3 I shall prove, using the formula

given in § 2, the single integral formula for $E[\mathrm{sgn}\,(x_1x_2x_3x_4)]$ originally due to Schläfli (1858, 1860), in the case of the general correlation matrix, and then shall give $E[|x_1x_2x_3x_4|]$ an exact expression involving $E[\mathrm{sgn}\,(x_1x_2x_3x_4)]$. Section 4 will be devoted to the absolute moments $E[|x_1^{n_1}x_2^{n_2}x_3^{n_3}x_4^{n_4}|]$ for the cases $2 = n_1 \geqslant n_2 \geqslant n_3 \geqslant n_4 \geqslant 1$. Finally in § 5, using the results given in § 3, I shall give the fourth moment of the mean deviation and of Gini's mean difference for a sample taken from a normal population.

2. *$E[x_1^{n_1}\dots x_r^{n_r}\mathrm{sgn}\,(x_{j_1}\dots x_{j_p})]$ and the incomplete moments.* Let $x_1, \dots, x_r$ be distributed according to an r-variate distribution, such that the absolute moments $E[|x_1^{m_1}\dots x_r^{m_r}|]$ are finite for any set of integers $m_1, \dots, m_r$ for which $0 \leqslant m_1 \leqslant n_1, \dots, 0 \leqslant m_r \leqslant n_r$, for a fixed set of non-negative integers $n_1, \dots, n_r$. Put $n = \sum_{j=1}^{r} n_j$. Let $\{j_1, \dots, j_p\}$ be an arbitrary subset of integers of the set $\{1, 2, \dots, r\}$, and $\{k_1, \dots, k_q\}$ be its complementary. Clearly $p+q = r$. Then by a similar argument as given in Nabeya (1952), we can prove

$$E[x_1^{n_1}\dots x_r^{n_r}\mathrm{sgn}\,(x_{j_1}\dots x_{j_p})] = E[x_1^{n_1}\dots x_r^{n_r}\mathrm{sgn}\,(x_{j_1})\dots\mathrm{sgn}\,(x_{j_p})]$$

$$= \frac{1}{i^{p+n}\pi^p}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\frac{dt_{j_1}\dots dt_{j_p}}{t_{j_1}\dots t_{j_p}}\left[\frac{\partial^n\phi(t_1, \dots, t_r)}{\partial t_1^{n_1}\dots\partial t_r^{n_r}}\right]_{t_{k_1}=\dots=t_{k_q}=0}, \tag{1}$$

where $\phi(t_1, \dots, t_r)$ denotes the characteristic function of $x_1, \dots, x_r$, and the integral in (1) with respect to $t_{j_1}, \dots, t_{j_p}$ is to mean Cauchy's principal value,

$$\lim_{\substack{\epsilon_1,\dots,\epsilon_p\to+0\\ C_1,\dots,C_p\to+\infty}}\left(\int_{\epsilon_1}^{C_1}+\int_{-C_1}^{-\epsilon_1}\right)dt_{j_1}\dots\left(\int_{\epsilon_p}^{C_p}+\int_{-C_p}^{-\epsilon_p}\right)dt_{j_p}.$$

If $p = 0$, (1) reduces to the well-known formula for obtaining the ordinary moments $E[x_1^{n_1}\dots x_r^{n_r}]$, and if $n_{j_1}, \dots, n_{j_p}$ are odd and $n_{k_1}, \dots, n_{k_q}$ are even, (1) is nothing else than the formula for absolute moments, $E[|x_1^{n_1}\dots x_r^{n_r}|]$, given in Nabeya (1952).

Now, let us consider the incomplete moment,

$$\int_0^{\infty}\dots\int_0^{\infty}x_1^{n_1}\dots x_r^{n_r}f(x_1, \dots, x_r)\,dx_1\dots dx_r, \tag{2}$$

where $f(x_1, \dots, x_r)$ is the density function of $x_1, \dots, x_r$. For this purpose we form the expression $\{1+\mathrm{sgn}\,(x_1)\}\dots\{1+\mathrm{sgn}\,(x_r)\}$. It is equal to 2^r when $x_1 > 0, \dots, x_r > 0$, but it is equal to zero elsewhere except for a set of Lebesgue measure zero. Therefore, (2) is equal to

$$\frac{1}{2^r}E[x_1^{n_1}\dots x_r^{n_r}\{1+\mathrm{sgn}\,(x_1)\}\dots\{1+\mathrm{sgn}\,(x_r)\}]$$

$$= \frac{1}{2^r}\Sigma E[x_1^{n_1}\dots x_r^{n_r}\mathrm{sgn}\,(x_{j_1}\dots x_{j_p})], \tag{3}$$

where the summation runs over 2^r terms, $\{j_1, \dots, j_p\}$ being any subset of $\{1, 2, \dots, r\}$.

Here we add two remarks. First, if f is an even function, i.e. $f(x_1, \dots, x_r) = f(-x_1, \dots, -x_r)$, as in the cases of the multivariate normal distribution with zero means, then the expectation (1) is equal to zero when $n+p$ is an odd number. Hence in these cases the summation (3) includes only 2^{r-1} terms. Secondly, in the general case, if we are interested in the contribution to the absolute moment from other orthants, we can use

$$\pm\{1\pm\mathrm{sgn}\,(x_1)\}\dots\{1\pm\mathrm{sgn}\,(x_r)\}, \quad \text{instead of} \quad \{1+\mathrm{sgn}\,(x_1)\}\dots\{1+\mathrm{sgn}\,(x_r)\},$$

with appropriate signs.

Example 1. Let $f(x_1, x_2, x_3)$ be the normal density function of x_1, x_2, x_3, with means zero, with variances unity, and with correlations ρ_{jk}. Find the value of the incomplete moment

$$\iiint_0^\infty x_1^2 x_2 x_3 f(x_1, x_2, x_3)\, dx_1 dx_2 dx_3.$$

In this case, since $r = 3$, and n is an even number, we need only the expectations of type (1) for which p is 0 or 2. They are

$$E[x_1^2 x_2 x_3] = \rho_{23} + 2\rho_{12}\rho_{13},$$

$$E[x_1^2 x_2 x_3 \operatorname{sgn}(x_1 x_2)] = \frac{2}{\pi}\{\sqrt{(1-\rho_{12}^2)}\,(2\rho_{13}+\rho_{12}\rho_{23}) + (\rho_{23}+2\rho_{12}\rho_{13})\sin^{-1}\rho_{12}\},$$

$$E[x_1^2 x_2 x_3 \operatorname{sgn}(x_1 x_3)] = \frac{2}{\pi}\{\sqrt{(1-\rho_{13}^2)}\,(2\rho_{12}+\rho_{13}\rho_{23}) + (\rho_{23}+2\rho_{12}\rho_{13})\sin^{-1}\rho_{13}\},$$

$$E[x_1^2 x_2 x_3 \operatorname{sgn}(x_2 x_3)] = \frac{2}{\pi}\{\sqrt{(1-\rho_{23}^2)}\,(1+\rho_{12}^2+\rho_{13}^2) + (\rho_{23}+2\rho_{12}\rho_{13})\sin^{-1}\rho_{23}\},$$

therefore, according to (3), the required value is

$$\frac{1}{4\pi}\{\sqrt{(1-\rho_{12}^2)}\,(2\rho_{13}+\rho_{12}\rho_{23}) + \sqrt{(1-\rho_{13}^2)}\,(2\rho_{12}+\rho_{13}\rho_{23}) + \sqrt{(1-\rho_{23}^2)}\,(1+\rho_{12}^2+\rho_{13}^2) + (\rho_{23}+2\rho_{12}\rho_{13})(\tfrac{1}{2}\pi + \sin^{-1}\rho_{12} + \sin^{-1}\rho_{13} + \sin^{-1}\rho_{23})\},$$

in accordance with the result given in Kamat (1958), which includes all the incomplete moments for $n \leqslant 4$ in the trivariate case.

Example 2. Let x_1, x_2, x_3, x_4 be distributed according to the four-variate normal distribution, with zero means, arbitrary variances, and with correlations ρ_{jk}. Find the probability $P(x_1 > 0, x_2 > 0, x_3 > 0, x_4 > 0)$.

The required value is an incomplete moment (2), putting $r = 4$ and $n = 0$. From (3) we have

$$P(x_1 > 0, x_2 > 0, x_3 > 0, x_4 > 0) = \frac{1}{16} + \frac{1}{8\pi}\sum_{1 \leqslant j < k \leqslant 4} \sin^{-1}\rho_{jk} + \tfrac{1}{16}E[\operatorname{sgn}(x_1 x_2 x_3 x_4)],$$

where $E[\operatorname{sgn}(x_j x_k)] = (2/\pi)\sin^{-1}\rho_{jk}$ is given in Kendall (1948), and $E[\operatorname{sgn}(x_1 x_2 x_3 x_4)]$ will be treated in the next section.

It may be seen from the above examples, that in the normal case the expectations of the type (1) have in general simpler expressions than incomplete moments. Hence it may occur that in evaluating incomplete moments the above method is simpler than Kamat's.

Expectations of type (1) in the general case are used, e.g. in evaluating the moments of the statistic $c_1 x_1 + c_2 x_2 + c_3 |x_3| + c_4 |x_4|$, which is an extended form of the ones treated by Kamat (1953*a*).

3. *$E[\operatorname{sgn}(x_1 x_2 x_3 x_4)]$ and $E[|x_1 x_2 x_3 x_4|]$.* Suppose that x_1, x_2, x_3, x_4 are distributed according to the four-variate normal distribution, with means zero, variances unity, and correlation matrix $\mathbf{R} = (\rho_{jk})$. According to (1), $E[\operatorname{sgn}(x_1 x_2 x_3 x_4)]$ is given by

$$E[\operatorname{sgn}(x_1 x_2 x_3 x_4)] = \frac{1}{\pi^4}\iiint\int_{-\infty}^{\infty} \phi(t_1, t_2, t_3, t_4)\frac{dt_1 dt_2 dt_3 dt_4}{t_1 t_2 t_3 t_4}, \tag{4}$$

where $\phi(t_1, t_2, t_3, t_4) = e^{-\frac{1}{2}\mathbf{t}'\mathbf{R}\mathbf{t}}$, and $\mathbf{t}$ is the column vector of t's. By applying Plackett's (1954) technique to the characteristic function, we can get from the right side of (4) the Schläfli integral (1858, 1860) for $E[\mathrm{sgn}\,(x_1x_2x_3x_4)]$. For example, let

$$\mathbf{K}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \rho_{23} & \rho_{24} \\ 0 & \rho_{32} & 1 & \rho_{34} \\ 0 & \rho_{42} & \rho_{43} & 1 \end{pmatrix}, \quad \mathbf{K}_2 = \begin{pmatrix} 1 & \rho_{12} & 0 & 0 \\ \rho_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho_{34} \\ 0 & 0 & \rho_{43} & 1 \end{pmatrix}, \quad \mathbf{K}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

then $\alpha\mathbf{R} + (1-\alpha)\,\mathbf{K}_i$ $(i = 1, 2, 3)$ is also a correlation matrix, if $0 \leqslant \alpha \leqslant 1$. Let

$$F_i(\alpha) = \frac{1}{\pi^4}\int\int\int\int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}\mathbf{t}'\{\alpha\mathbf{R} + (1-\alpha)\,\mathbf{K}_i\}\,\mathbf{t}\right]\frac{dt_1dt_2dt_3dt_4}{t_1t_2t_3t_4},$$

then $\{\partial F_i(\alpha)\}/\partial\alpha$ can be expressed by means of algebraic and arcsine functions of ρ_{jk}'s. Let us show this for $\{\partial F_1(\alpha)\}/\partial\alpha$. We have

$$\begin{aligned}\frac{\partial F_1(\alpha)}{\partial\alpha} &= -\frac{1}{\pi^4}\sum_{j=2}^{4}\rho_{1j}\int\int\int\int_{-\infty}^{\infty}\exp\left[-\tfrac{1}{2}\mathbf{t}'\{\alpha\mathbf{R} + (1-\alpha)\,\mathbf{K}_i\}\,\mathbf{t}\right]\frac{t_j dt_1dt_2dt_3dt_4}{t_2t_3t_4} \\ &= -\frac{1}{\pi^4}\sum_{j=2}^{4}\rho_{1j}G_j(\alpha),\end{aligned}$$

say, where $G_2(\alpha)$, $G_3(\alpha)$, $G_4(\alpha)$ have similar expressions. To calculate $G_2(\alpha)$, carrying out the integration, first, with respect to t_1 and t_2, then with respect to t_3 and t_4, we have

$$\begin{aligned}G_2(\alpha) &= \frac{2\pi}{\sqrt{(1-\alpha^2\rho_{12}^2)}}\int\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2(1-\alpha^2\rho_{12}^2)}\{R_{44}^{(1)}(\alpha)\,t_3^2 - 2R_{34}^{(1)}(\alpha)\,t_3t_4 + R_{33}^{(1)}(\alpha)\,t_4^2\}\right]\frac{dt_3dt_4}{t_3t_4} \\ &= -\frac{4\pi^2}{\sqrt{(1-\alpha^2\rho_{12}^2)}}\sin^{-1}\rho_{34\cdot12}^{(1)}(\alpha),\end{aligned}$$

where $R_{jk}^{(i)}(\alpha)$ is the cofactor of (j,k) element in the determinant of $\alpha\mathbf{R} + (1-\alpha)\,\mathbf{K}_i$, and $\rho_{lm\cdot jk}^{(i)}(\alpha)$ is the partial correlation coefficient, given that the correlation matrix is $\alpha\mathbf{R} + (1-\alpha)\,\mathbf{K}_i$. Similarly for $G_3(\alpha)$ and $G_4(\alpha)$. Similar expressions hold for $\{\partial F_2(\alpha)\}/\partial\alpha$ and $\{\partial F_3(\alpha)\}/\partial\alpha$, so we have the Schläfli integral

$$\begin{aligned}E[\mathrm{sgn}\,(x_1x_2x_3x_4)] &= F_i(1) = F_i(0) + \int_0^1\frac{\partial F_i(\alpha)}{\partial\alpha}\,d\alpha \\ &= \frac{4}{\pi^2}\int_0^1\left\{\frac{\rho_{12}}{\sqrt{(1-\alpha^2\rho_{12}^2)}}\sin^{-1}\rho_{34\cdot12}^{(1)}(\alpha) + \frac{\rho_{13}}{\sqrt{(1-\alpha^2\rho_{13}^2)}}\sin^{-1}\rho_{24\cdot13}^{(1)}(\alpha)\right. \\ &\qquad \left. + \frac{\rho_{14}}{\sqrt{(1-\alpha^2\rho_{14}^2)}}\sin^{-1}\rho_{23\cdot14}^{(1)}(\alpha)\right\}d\alpha, \qquad (5) \\ &= \frac{4}{\pi^2}\left[\sin^{-1}\rho_{12}\sin^{-1}\rho_{34} + \int_0^1\left\{\frac{\rho_{13}}{\sqrt{(1-\alpha^2\rho_{13}^2)}}\sin^{-1}\rho_{24\cdot13}^{(2)}(\alpha) + \frac{\rho_{14}}{\sqrt{(1-\alpha^2\rho_{14}^2)}}\sin^{-1}\rho_{23\cdot14}^{(2)}(\alpha)\right.\right. \\ &\qquad \left.\left. + \frac{\rho_{23}}{\sqrt{(1-\alpha^2\rho_{23}^2)}}\sin^{-1}\rho_{14\cdot23}^{(2)}(\alpha) + \frac{\rho_{24}}{\sqrt{(1-\alpha^2\rho_{24}^2)}}\sin^{-1}\rho_{13\cdot24}^{(2)}(\alpha)\right\}d\alpha\right], \qquad (6) \\ &= \frac{4}{\pi^2}\int_0^1\Sigma\frac{\rho_{jk}}{\sqrt{(1-\alpha^2\rho_{jk}^2)}}\sin^{-1}\rho_{lm\cdot jk}^{(3)}(\alpha)\,d\alpha, \qquad (7)\end{aligned}$$

where the summation in (7) runs over six terms for which (j,k,l,m) is a permutation of $(1,2,3,4)$ such that $j < k$ and $l < m$. These formulae can be regarded as equivalent to the special expressions of $\Phi_4^0(P)$ given by Plackett (1954), but I believe that the above choice of K's would provide a better basis for the calculation than his choice.

If $\rho_{jk} = \rho$ $(j \neq k)$, then the above formulae reduce to

$$E[\text{sgn}(x_1x_2x_3x_4)] = \frac{12}{\pi^2}\int_0^\rho \frac{1}{\sqrt{(1-\alpha^2)}}\sin^{-1}\frac{\rho-\alpha^2}{1+\rho-2\alpha^2}d\alpha, \tag{8}$$

$$= \frac{4}{\pi^2}(\sin^{-1}\rho)^2 + \frac{16}{\pi^2}\int_0^\rho \frac{1}{\sqrt{(1-\alpha^2)}}\sin^{-1}\frac{(1-\rho)\alpha}{1+\rho-2\alpha^2}d\alpha, \tag{9}$$

$$= \frac{24}{\pi^2}\int_0^\rho \frac{1}{\sqrt{(1-\alpha^2)}}\sin^{-1}\frac{\alpha}{1+2\alpha}d\alpha. \tag{10}$$

Now we proceed to evaluate $E[|x_1x_2x_3x_4|]$ under the same assumption as above. It is expressed by (1) as

$$E[|x_1x_2x_3x_4|] = \frac{1}{\pi^4}\int\int\int\int_{-\infty}^{\infty} \frac{\partial^4\phi(t_1,t_2,t_3,t_4)}{\partial t_1\partial t_2\partial t_3\partial t_4}\frac{dt_1dt_2dt_3dt_4}{t_1t_2t_3t_4}. \tag{11}$$

The partial derivative appearing in (11) is of the form

$$\frac{\partial^4\phi(t_1,t_2,t_3,t_4)}{\partial t_1\partial t_2\partial t_3\partial t_4} = \phi(t_1,t_2,t_3,t_4)\times(\text{polynomial in } t\text{'s of the fourth degree});$$

therefore, in evaluating (11) we must find

$$H(\gamma_1,\gamma_2,\gamma_3,\gamma_4) = \frac{1}{\pi^4}\int\int\int\int_{-\infty}^{\infty}\phi(t_1,t_2,t_3,t_4)\frac{t_1^{\gamma_1}t_2^{\gamma_2}t_3^{\gamma_3}t_4^{\gamma_4}}{t_1t_2t_3t_4}dt_1dt_2dt_3dt_4$$

for all sets of non-negative integers $\gamma_1, \gamma_2, \gamma_3$ and γ_4, such that $\gamma_1+\gamma_2+\gamma_3+\gamma_4 = 0, 2$, or 4. According to (4) we have $H(0,0,0,0) = E[\text{sgn}(x_1x_2x_3x_4)]$, and the other H's may be obtained by the same method as shown in Nabeya (1952). Thus we have

$$\begin{aligned}E[|x_1x_2x_3x_4|] = \frac{4}{\pi^2}\{&\sqrt{R}+\sqrt{(1-\rho_{12}^2)}(\rho_{34}+\rho_{13}\rho_{14}+\rho_{23}\rho_{24})\sin^{-1}\rho_{34\cdot12}\\ &+\sqrt{(1-\rho_{13}^2)}(\rho_{24}+\rho_{12}\rho_{14}+\rho_{23}\rho_{34})\sin^{-1}\rho_{24\cdot13}+\sqrt{(1-\rho_{14}^2)}(\rho_{23}+\rho_{12}\rho_{13}+\rho_{24}\rho_{34})\sin^{-1}\rho_{23\cdot14}\\ &+\sqrt{(1-\rho_{23}^2)}(\rho_{14}+\rho_{12}\rho_{24}+\rho_{13}\rho_{34})\sin^{-1}\rho_{14\cdot23}+\sqrt{(1-\rho_{24}^2)}(\rho_{13}+\rho_{12}\rho_{23}+\rho_{14}\rho_{34})\sin^{-1}\rho_{13\cdot24}\\ &+\sqrt{(1-\rho_{34}^2)}(\rho_{12}+\rho_{13}\rho_{23}+\rho_{14}\rho_{24})\sin^{-1}\rho_{12\cdot34}\}\\ &+(\rho_{12}\rho_{34}+\rho_{13}\rho_{24}+\rho_{14}\rho_{23})E[\text{sgn}(x_1x_2x_3x_4)],\end{aligned} \tag{12}$$

where R denotes the determinant of $\mathbf{R}$.

The formula checks with the previous ones in all the cases where $\rho_{12} = \rho_{13} = \rho_{14} = 0$, or $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$, or $\rho_{14} = 1$, $\rho_{12} = \rho_{24}$ and $\rho_{13} = \rho_{34}$. In the case $\rho_{jk} = \rho$ $(j \neq k)$, we have

$$\begin{aligned}E[|x_1x_2x_3x_4|] = \frac{4}{\pi^2}&\left\{\sqrt{[(1+3\rho)(1-\rho)^3]}+6\rho(1+2\rho)\sqrt{(1-\rho^2)}\sin^{-1}\frac{\rho}{1+2\rho}\right\}\\ &+3\rho^2E[\text{sgn}(x_1x_2x_3x_4)],\end{aligned} \tag{13}$$

where $E[\text{sgn}(x_1x_2x_3x_4)]$ was given in (8)–(10).

4. $E[|x_1^{n_1}x_2^{n_2}x_3^{n_3}x_4^{n_4}|]$ *for cases* $2 = n_1 \geqslant n_2 \geqslant n_3 \geqslant n_4 \geqslant 1$. In the following we give the absolute moments $E[|x_1^{n_1}x_2^{n_2}x_3^{n_3}x_4^{n_4}|]$ for the cases $2 = n_1 \geqslant n_2 \geqslant n_3 \geqslant n_4 \geqslant 1$. In evaluating these we have no more difficulty than in the two- or three-variate cases.

$$\begin{aligned}E[|x_1^2x_2x_3x_4|] = (2/\pi)^3\{&\sqrt{(R_{11})}(1+\rho_{12}^2+\rho_{13}^2+\rho_{14}^2)\\ &+(\rho_{34}+2\rho_{13}\rho_{14}+\rho_{23}\rho_{24}+2\rho_{12}\rho_{13}\rho_{24}+2\rho_{12}\rho_{14}\rho_{23}+\rho_{12}^2\rho_{34}-\rho_{12}^2\rho_{23}\rho_{24})\sin^{-1}\rho_{34\cdot2}\\ &+(\rho_{24}+2\rho_{12}\rho_{14}+\rho_{23}\rho_{34}+2\rho_{12}\rho_{13}\rho_{34}+2\rho_{13}\rho_{14}\rho_{23}+\rho_{13}^2\rho_{24}-\rho_{13}^2\rho_{23}\rho_{34})\sin^{-1}\rho_{24\cdot3}\\ &+(\rho_{23}+2\rho_{12}\rho_{13}+\rho_{24}\rho_{34}+2\rho_{12}\rho_{14}\rho_{34}+2\rho_{13}\rho_{14}\rho_{24}+\rho_{14}^2\rho_{23}-\rho_{14}^2\rho_{24}\rho_{34})\sin^{-1}\rho_{23\cdot4}\},\end{aligned}$$

$$E[|x_1^2x_2^2x_3x_4|] = (2/\pi)\{\sqrt{(1-\rho_{34}^2)}\,[2\rho_{12}^2+2\rho_{13}^2+2\rho_{14}^2+2\rho_{23}^2+2\rho_{24}^2+\rho_{34}^2$$
$$+4\rho_{12}\rho_{13}\rho_{23}+4\rho_{12}\rho_{14}\rho_{24}-2\rho_{13}\rho_{14}\rho_{34}-2\rho_{23}\rho_{24}\rho_{34}-2\rho_{13}^2\rho_{23}^2-2\rho_{14}^2\rho_{24}^2$$
$$+4\rho_{13}\rho_{14}\rho_{23}\rho_{24}+R_{11}R_{22}/(1-\rho_{34}^2)]+(\rho_{34}+2\rho_{13}\rho_{14}+2\rho_{23}\rho_{24}+4\rho_{12}\rho_{13}\rho_{24}$$
$$+4\rho_{12}\rho_{14}\rho_{23}+2\rho_{12}^2\rho_{34})\sin^{-1}\rho_{34}\},$$

$$E[|x_1^2x_2^2x_3^2x_4|] = (2/\pi)^{\frac{1}{2}}(1+2\rho_{12}^2+2\rho_{13}^2+2\rho_{23}^2+\rho_{14}^2+\rho_{24}^2+\rho_{34}^2$$
$$+8\rho_{12}\rho_{13}\rho_{23}+4\rho_{12}\rho_{14}\rho_{24}+4\rho_{13}\rho_{14}\rho_{34}+4\rho_{23}\rho_{24}\rho_{34}+2\rho_{12}^2\rho_{34}^2$$
$$+2\rho_{13}^2\rho_{24}^2+2\rho_{14}^2\rho_{23}^2-\rho_{14}^2\rho_{24}^2-\rho_{14}^2\rho_{34}^2-\rho_{24}^2\rho_{34}^2+8\rho_{12}\rho_{13}\rho_{24}\rho_{34}$$
$$+8\rho_{12}\rho_{14}\rho_{23}\rho_{34}+8\rho_{13}\rho_{14}\rho_{23}\rho_{24}-4\rho_{12}\rho_{14}\rho_{24}\rho_{34}^2-4\rho_{13}\rho_{14}\rho_{24}^2\rho_{34}$$
$$-4\rho_{14}^2\rho_{23}\rho_{24}\rho_{34}+3\rho_{14}^2\rho_{24}^2\rho_{34}^2),$$

$$E[x_1^2x_2^2x_3^2x_4^2] = 1+2\rho_{12}^2+2\rho_{13}^2+2\rho_{14}^2+2\rho_{23}^2+2\rho_{24}^2+2\rho_{34}^2$$
$$+8\rho_{12}\rho_{13}\rho_{23}+8\rho_{12}\rho_{14}\rho_{24}+8\rho_{13}\rho_{14}\rho_{34}+8\rho_{23}\rho_{24}\rho_{34}+4\rho_{12}^2\rho_{34}^2+4\rho_{13}^2\rho_{24}^2$$
$$+4\rho_{14}^2\rho_{23}^2+16\rho_{12}\rho_{13}\rho_{24}\rho_{34}+16\rho_{12}\rho_{14}\rho_{23}\rho_{34}+16\rho_{13}\rho_{14}\rho_{23}\rho_{24}.$$

5. *The fourth moment of mean deviation and of Gini's mean difference.* Let $z_1, \ldots, z_n$ be a sample from a normal population with variance σ^2, and let $\bar{z}$ denote their mean. Then

$$m = \frac{1}{n}\sum_{j=1}^{n}|z_j-\bar{z}| \quad \text{and} \quad g = \frac{2}{n(n-1)}\sum_{1\leqslant j<k\leqslant n}|z_j-z_k|$$

are the mean deviation and Gini's mean difference for the sample, respectively. In this section we evaluate the fourth moments $\mu_4'(m)$ and $\mu_4'(g)$ of these two statistics.

First, we consider $\mu_4'(m)$. Let $x_j = z_j-\bar{z}$ $(1 \leqslant j \leqslant n)$; then the only difficulty in evaluating $\mu_4'(m)$ is in finding the exact expression for $E[|x_1x_2x_3x_4|]$. It is given by $(n-1)^2\sigma^4/n^2$ times (13), putting $\rho = -1/(n-1)$, and we have

$$\mu_4'(m) = \frac{(n-1)^2}{n^5}\sigma^4\Big[3+(n-1)\Big\{3+6\rho^2+\frac{8}{\pi}\sqrt{(1-\rho^2)}\,(2+\rho^2)+\frac{24}{\pi}\rho\sin^{-1}\rho\Big\}$$
$$+\frac{12}{\pi}(n-1)^{[2]}\{\sqrt{(1-\rho^2)}\,(1+2\rho^2)+\rho(1+2\rho)\sin^{-1}\rho\}$$
$$+(n-1)^{[3]}\Big\{\frac{4}{\pi^2}\sqrt{\{(1+3\rho)\,(1-\rho)^3\}}+\frac{24}{\pi^2}\rho(1+2\rho)\sqrt{(1-\rho^2)}\sin^{-1}\frac{\rho}{1+2\rho}$$
$$+3\rho^2E[\operatorname{sgn}(x_1x_2x_3x_4)]\Big\}\Big].$$

Geary (1936) has given $E[|x_1x_2x_3x_4|]$ the form of a series expansion in powers of $(n-2)^{-1}$, namely

$$E[|x_1x_2x_3x_4|] = \frac{4}{\pi^2}\frac{(n+1)^{\frac{1}{2}}(n-3)^{\frac{7}{2}}}{(n-2)^4}\sigma^4\Big\{1+\frac{12}{(n-2)^2}-\frac{32}{(n-2)^3}$$
$$+\frac{172}{(n-2)^4}-\frac{640}{(n-2)^5}+\frac{12736}{5(n-2)^6}-\frac{47104}{5(n-2)^7}+\ldots\Big\}. \qquad (14)$$

Both (14) and $(n-1)^2\sigma^4/n^2$ times (13), if they are written in the form of series expansion in $\rho = -(n-1)^{-1}$, agree with

$$E[|x_1x_2x_3x_4|] = \frac{4}{\pi^2}\left(\frac{n-1}{n}\right)^2\sigma^4(1+3\rho^2+4\rho^3+\rho^4-4\rho^5+\tfrac{31}{5}\rho^6-\tfrac{48}{5}\rho^7\ldots).$$

Now we proceed to the evaluation of $\mu_4'(g)$. Let $x_{jk} = z_j - z_k$, then we have

$$\begin{aligned}\mu_4'(g) = \frac{16}{n^4(n-1)^4}\{&\tfrac{1}{2}n^{[2]}E[x_{12}^4] + 4n^{[3]}E[|x_{12}^3 x_{13}|] \\ &+ n^{[4]}E[|x_{12}^3 x_{34}|] + 3n^{[3]}E[x_{12}^2 x_{13}^2] + \tfrac{3}{4}n^{[4]}E[x_{12}^2 x_{34}^2] \\ &+ 6n^{[4]}E[|x_{12}^2 x_{13} x_{14}|] + 12n^{[4]}E[|x_{12}^2 x_{13} x_{34}|] + 6n^{[3]}E[|x_{12}^2 x_{13} x_{23}|] \\ &+ 6n^{[4]}E[|x_{12}^2 x_{13} x_{24}|] + 6n^{[5]}E[|x_{12}^2 x_{13} x_{45}|] + 3n^{[5]}E[|x_{12}^2 x_{34} x_{35}|] \\ &+ \tfrac{3}{4}n^{[6]}E[|x_{12}^2 x_{34} x_{56}|] + n^{[5]}E[|x_{12} x_{13} x_{14} x_{15}|] + 12n^{[4]}E[|x_{12} x_{13} x_{14} x_{23}|] \\ &+ 12n^{[5]}E[|x_{12} x_{13} x_{14} x_{25}|] + 2n^{[6]}E[|x_{12} x_{13} x_{14} x_{56}|] + 3n^{[4]}E[|x_{12} x_{13} x_{24} x_{34}|] \\ &+ 12n^{[5]}E[|x_{12} x_{13} x_{24} x_{35}|] + 2n^{[5]}E[|x_{12} x_{13} x_{23} x_{45}|] + 6n^{[6]}E[|x_{12} x_{13} x_{24} x_{56}|] \\ &+ 3n^{[6]}E[|x_{12} x_{13} x_{45} x_{46}|] + \tfrac{3}{2}n^{[7]}E[|x_{12} x_{13} x_{45} x_{67}|] + \tfrac{1}{16}n^{[8]}E[|x_{12} x_{34} x_{56} x_{78}|]\}.\end{aligned}$$

Among the twenty-three expectations appearing in the above expressions, eighteen may be evaluated by means of the formulae already published. To evaluate the remaining five, (12) is useful. Let us, for example, consider $E[|x_{12}x_{13}x_{14}x_{25}|]$. We have

$$\begin{gathered}\sigma(x_{12}) = \sigma(x_{13}) = \sigma(x_{14}) = \sigma(x_{25}) = \sqrt{2}\,\sigma, \\ \rho(x_{12}, x_{13}) = \rho(x_{12}, x_{14}) = \rho(x_{13}, x_{14}) = \tfrac{1}{2}, \\ \rho(x_{12}, x_{25}) = -\tfrac{1}{2}, \quad \rho(x_{13}, x_{25}) = \rho(x_{14}, x_{25}) = 0,\end{gathered}$$

and $\operatorname{sgn}(x_{12}x_{13}x_{14}x_{25})$ is determined by the order of magnitude of z_1, z_2, z_3, z_4 and z_5. In this case every order of z's is equally probable, so that $E[\operatorname{sgn}(x_{12}x_{13}x_{14}x_{25})] = -\frac{1}{15}$. Hence we get

$$E[|x_{12}x_{13}x_{14}x_{25}|] = \left\{\frac{1}{\pi^2}\left(4\sqrt{5} + 2\sqrt{3}\sin^{-1}\tfrac{1}{4} + 24\sin^{-1}\frac{1}{\sqrt{6}} + 4\sqrt{3}\sin^{-1}\sqrt{\tfrac{3}{8}}\right) + \tfrac{1}{15}\right\}\sigma^4.$$

We can obtain similarly

$$E[|x_{12}x_{13}x_{14}x_{15}|] = \left\{\frac{1}{\pi^2}(4\sqrt{5} + 48\sqrt{3}\sin^{-1}\tfrac{1}{4}) + \tfrac{3}{5}\right\}\sigma^4,$$

$$E[|x_{12}x_{13}x_{14}x_{23}|] = \frac{1}{\pi}(4 + 2\sqrt{3})\,\sigma^4,$$

$$E[|x_{12}x_{13}x_{24}x_{34}|] = \left\{\frac{1}{\pi}(8\sqrt{3} - 8) + \tfrac{2}{3}\right\}\sigma^4,$$

$$E[|x_{12}x_{13}x_{24}x_{35}|] = \left\{\frac{1}{\pi^2}\left(4\sqrt{5} + 8\sin^{-1}\tfrac{2}{3} - 8\sin^{-1}\frac{1}{\sqrt{6}} + 8\sqrt{3}\sin^{-1}\sqrt{\tfrac{3}{8}}\right) + \tfrac{2}{15}\right\}\sigma^4.$$

The last expectation denoted by $\mathscr{E}(d_1d_2d_3d_4)$ has been obtained by Kamat (1953*b*), numerically, using my integral formula (Nabeya, 1952). Finally we have

$$\begin{aligned}\mu_4'(g) = \frac{16\sigma^4}{n^3(n-1)^3}\Bigg\{&6 + \frac{n-2}{\pi}(72\sqrt{3} + 26\pi) + \frac{(n-2)^{[2]}}{\pi}(112 + 144\sqrt{3} + 21\pi) \\ &+ \frac{(n-2)^{[3]}}{\pi^2}\left(100\sqrt{5} + 72\sqrt{3}\sin^{-1}\tfrac{1}{4} + 96\sin^{-1}\tfrac{2}{3} + 192\sin^{-1}\frac{1}{\sqrt{6}} + 144\sqrt{3}\sin^{-1}\sqrt{\tfrac{3}{8}}\right. \\ &\left. + 72\pi + 12\sqrt{3}\,\pi + 5\pi^2\right) + \frac{(n-2)^{[4]}}{\pi^2}\left(36 + 64\sqrt{2} + 48\sin^{-1}\tfrac{1}{3} + 96\sin^{-1}\frac{1}{\sqrt{3}}\right. \\ &\left. + 6\pi + 4\sqrt{3}\,\pi + \tfrac{1}{3}\pi^2\right) + \frac{(n-2)^{[5]}}{\pi^2}(12\sqrt{3} + 2\pi) + \frac{(n-2)^{[6]}}{\pi^2}\Bigg\}.\end{aligned}$$

I express sincere thanks to Prof. E. S. Pearson and the referees for calling my attention to Plackett's and Schläfli's work.

REFERENCES

GEARY, R. C. (1936). Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biometrika*, **28**, 295–305.

KAMAT, A. R. (1953*a*). Incomplete and absolute moments of the multivariate normal distribution with some applications. *Biometrika*, **40**, 20–34.

KAMAT, A. R. (1953*b*). On the mean successive difference and its ratio to the root mean square. *Biometrika*, **40**, 116–27.

KAMAT, A. R. (1953*c*). The third moment of Gini's mean differences. *Biometrika*, **40**, 451–2.

KAMAT, A. R. (1954). Moments of a mean deviation. *Biometrika*, **41**, 541–2.

KAMAT, A. R. (1958). Incomplete moments of the trivariate normal distribution. *Sankhyā*, **20**, 321–2.

KENDALL, M. G. (1948). *Rank Correlation Methods*. London: C. Griffin and Co.

NABEYA, S. (1951). Absolute moments in 2-dimensional normal distribution. *Ann. Inst. Statist. Math.* **3**, 2–6.

NABEYA, S. (1952). Absolute moments in 3-dimensional normal distribution. *Ann. Inst. Statist. Math.* **4**, 15–29.

PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351–60.

SCHLÄFLI, L. (1858, 1860). On the multiple integral $\int^n dxdy \ldots dz$ whose limits are

$$p_1 = a_1x+b_1y+\ldots+h_1z > 0,\ p_2 > 0, \ldots, p_n > 0 \quad \text{and} \quad x^2+y^2+\ldots+z^2 < 1.$$

Quart. J. Pure Appl. Math. **2**, 269–301; **3**, 54–68, 97–108.

Biometrika (1961), **48**, 1 *and* 2, *p.* 85
Printed in Great Britain

Asymptotic expansions for the mean and variance of the serial correlation coefficient

BY JOHN S. WHITE
Mathematics Group, General Motors Research Laboratories

SUMMARY

Two models are considered for the estimation of the serial correlation coefficient α of a first-order auto-regressive Gaussian process. Series expansions are obtained for the first two moments of $\hat{\alpha}$, the least squares estimator for α. The series expansions are carried to terms of order T^{-3} and α^4 (where T is the observed length of the series) thus extending the asymptotic results of several authors (e.g. Bartlett, 1946; Hurwicz, 1950; Kendall, 1954; Marriott & Pope, 1954).

INTRODUCTION

We consider the discrete Gaussian process (x_t) satisfying the stochastic difference equation

$$x_t = \alpha x_{t-1} + u_t, \tag{1}$$

where α is an unknown parameter and the u's are independent normal variables with zero mean and variance σ^2. The distribution of a sample $x = (x_1, \ldots, x_T)$ is not uniquely determined by (1) but depends also upon the specification of an initial condition. If we take $x_0 = c$ (a constant) then the density function for x is

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}T} \exp\left\{\frac{-\sum_1^T (x_t - \alpha x_{t-1})^2}{2\sigma^2}\right\}. \tag{2}$$

A second possible initial condition is that the process be stationary. This implies that each x_t is $N\{0, \sigma^2/(1-\alpha^2)\}$ and that the density function for x is

$$f^*(x) = (2\pi\sigma^2)^{-\frac{1}{2}T}(1-\alpha^2)^{\frac{1}{2}} \exp\left\{\frac{-1}{2\sigma^2}\left[(1-\alpha^2)x_1^2 + \sum_1^T (x_t - \alpha x_{t-1})^2\right]\right\}. \tag{3}$$

We shall speak of the process which satisfies (2) as model A and the process which satisfies (3) as model B. To differentiate between the two models we shall frequently affix an asterisk to expressions referring to model B.

MAXIMUM LIKELIHOOD ESTIMATORS

For model A the maximum likelihood estimators are

$$\hat{\alpha} = \frac{\Sigma x_t x_{t-1}}{\Sigma x_{t-1}^2}, \quad \hat{\sigma}^2 = \frac{\Sigma (x_t - \hat{\alpha} x_{t-1})^2}{T}.$$

It may also be noted that $\hat{\alpha}$ is the least squares estimator as obtained by minimizing

$$\sum_1^T (x_t - \alpha x_{t-1})^2 = \Sigma u_t^2.$$

Following Koopmans (1942) we define the statistics

$$L = x_1^2 + x_T^2,$$
$$M = x_1x_2 + x_2x_3 + \ldots + x_{T-1}x_T,$$
$$N = x_2^2 + \ldots + x_{T-1}^2.$$

For model B the maximum likelihood estimators are then the solutions of the equations

$$(M - \alpha N)(1 - \alpha^2) - \frac{\alpha}{T}(L - 2\alpha M + N + \alpha^2 N) = 0,$$

$$\sigma^2 = \frac{L - 2\alpha M + N + \alpha^2 N}{T}.$$

The first of these equations may be written as

$$g(\alpha) = \alpha^3 - \frac{T-2}{T-1}\frac{M}{N}\alpha^2 - \frac{L + (T+1)N}{(T-1)N}\alpha + \frac{T}{T-1}\frac{M}{N} = 0.$$

The three roots of $g(\alpha)$ may be located by considering

$$g(+\infty) = +\infty,$$
$$g(1) = \frac{2M - 2N - L}{N(T-1)} < 0,$$
$$g(-1) = \frac{2M + 2N + L}{N(T-1)} > 0,$$
$$g(-\infty) = -\infty.$$

Thus $g(\alpha) = 0$ has three roots, say α^*, α_1^*, and α_2^* such that

$$\alpha_1^* < -1 < \alpha^* < 1 < \alpha_2^*$$

and the maximum likelihood estimator for α is α^*, the unique root between -1 and $+1$.

For large T we have

$$g(\alpha) \cong \alpha^3 - \frac{M}{N}\alpha^2 - \alpha + \frac{M}{N}$$
$$= \left(\alpha - \frac{M}{N}\right)(\alpha^2 - 1).$$

Hence, the three roots of $g(\alpha)$ are asymptotically 1, -1 and M/N.

For our two models we then have as estimators for α

$$\hat{\alpha} = \frac{x_1x_2 + x_2x_3 + \ldots + x_{T-1}x_T}{x_1^2 + x_2^2 + \ldots + x_{T-1}^2},$$
$$\frac{M}{N} = \frac{x_1x_2 + x_2x_3 + \ldots + x_{T-1}x_T}{x_2^2 + \ldots + x_{T-1}^2}.$$

Since these two estimators are asymptotically the same, we shall focus our attention on $\hat{\alpha}$ and consider some of its properties with respect to both models.

Model A, $x_0 = 0$

We shall consider model A first for the case $c = x_0 = 0$. Since $\hat{\alpha}$ is independent of σ^2 we may also set $\sigma^2 = 1$. Following the procedure used by Williams (1941) and Hurwicz (1950) we set $U = \Sigma x_t x_{t-1}$ and $V = \Sigma x_{t-1}^2$. The joint moment generating function of U and V is then

$$E(\exp\{Uu+Vv\}) = m(u,v) = \int \exp(Uu+Vv) f(x)\,dx$$

$$= (2\pi)^{-\frac{1}{2}T} \exp \int (-\tfrac{1}{2} x D x')\,dx = |D|^{-\frac{1}{2}}, \tag{4}$$

where D is the $T \times T$ matrix whose determinant is

$$|D| = D(T) = \begin{vmatrix} p & q & 0 & \dots & \dots & \dots \\ q & p & q & & & \\ 0 & q & p & & & \\ \dots & \dots & \dots & & & \\ & & & q & p & q \\ & & & 0 & q & 1 \end{vmatrix}, \tag{5}$$

$$p = 1+\alpha^2-2v, \quad q = -(\alpha+u).$$

The moments of $\hat{\alpha} = U/V$ may then be obtained from

$$E(\hat{\alpha}^k) = \int_{-\infty}^{0}\int_{-\infty}^{v_k} \dots \int_{-\infty}^{v_2} \frac{\partial^k m(u,v)}{\partial u^k}\bigg|_{u=0} dv_1 dv_2 \dots dv_k.$$

Hurwicz has noted that even for small values of T ($T \geqslant 6$) the integrals involved in evaluating $E(\hat{\alpha})$ are hyperelliptic and their evaluation in closed form is next to impossible. On the other hand it is feasible to expand the integrand in a Maclaurin series and integrate termwise.

It turns out that $E(\hat{\alpha})$ is an odd function of α while $E(\hat{\alpha}^2)$ is an even function of α. Setting $\beta = \alpha^2$, we define

$$\frac{E(\hat{\alpha})}{\alpha} = Q(\beta) = Q(0)+Q'(0)\beta+Q''(0)\frac{\beta^2}{2}+\dots,$$

$$E(\hat{\alpha}^2) = R(\beta) = R(0)+R'(0)+R''(0)\frac{\beta^2}{2}+\dots.$$

To expand the integrand of (8) in a Maclaurin series we note that, since $D(T)^{-\frac{1}{2}} = m(u,v)$, we have, for example,

$$Q(\beta) = \frac{E(\hat{\alpha})}{\alpha} = \frac{1}{\alpha}\int_{-\infty}^{0} \frac{\partial m(u,v)}{\partial u}\bigg|_{u=0} dv$$

$$= -\frac{1}{2}\int_{-\infty}^{0} \frac{1}{\alpha} D(T)^{-\frac{3}{2}} \frac{\partial D(T)}{\partial u}\bigg|_{u=0} dv, \tag{6}$$

$$Q'(0) = -\frac{1}{2}\int_{-\infty}^{0} \frac{\partial}{\partial\beta}\left(\frac{1}{\alpha} D(T)^{-\frac{3}{2}} \frac{\partial D(T)}{\partial u}\bigg|_{u=0}\right)_{\beta=0} dv,$$

with similar expressions for the remaining coefficients in $Q(\beta)$ and $R(\beta)$.

We note from (5), expanding $D(T)$ by elements of the first row, that $D(T)$ satisfies the difference equation

$$D(T) = pD(T-1) - q^2 D(T-2). \tag{7}$$

From the initial conditions $D(1) = 1$, $D(2) = p - q^2$ we find that

$$\begin{aligned} D(T) &= p^{T-2} \sum_{k=0} (-1)^k \left\{ p \binom{T-k-1}{k} - q^2 \binom{T-k-2}{k} \right\} p^{-2k} q^{2k} \\ &= \sum_{k=0} (-1)^k \left\{ (x+\beta) \binom{T-k-1}{k} - y \binom{T-k-2}{k} \right\} (x+\beta)^{T-2-2k} y^k \\ &\qquad (x = 1-2v, y = (\alpha+u)^2). \end{aligned} \tag{8}$$

The various derivatives required to evaluate expressions such as (10) may now be obtained from (8). Writing

$$D(T,i,j) = \frac{\partial}{\partial \beta^j} \left(\frac{\partial}{\partial y^i} D(T) \Big|_{y=\beta} \right) \Big|_{\beta=0},$$

$$D(T,0,0) = x^{T-1},$$

$$D(T,0,1) = (T-2) x^{T-3} (x-1),$$

$$D(T,0,2) = (T-3) x^{T-5} \{ (T-2) x^2 - 2(T-3) x + (T-4) \},$$

$$D(T,1,0) = -x^{T-3} (x + T - 2),$$

$$D(T,1,1) = -x^{T-5} \{ (T-2) x^2 + (T-3)(T-4) x - (T-3)(T-4) \},$$

$$D(T,1,2) = -x^{T-7} \{ (T-2)(T-3) x^3 + (T-3)(T-4)(T-6) x^2 - (T-4)(T-5)(2T-9) x + (T-4)(T-5)(T-6) \},$$

$$D(T,2,0) = (T-3) x^{T-5} (2x + T - 4),$$

$$D(T,2,1) = (T-4) x^{T-7} \{ 2(T-3) x^2 + (T-5)(T-6) x - (T-5)(T-6) \}.$$

As an example we evaluate $Q'(0)$ explicitly. We have

$$\begin{aligned} \frac{1}{\alpha} \frac{\partial D(T)}{\partial u} \Big|_{u=0} &= \frac{1}{\alpha} \frac{\partial D(T)}{\partial y} \frac{\partial y}{\partial u} \Big|_{u=0} \\ &= \frac{2(\alpha+u)}{\alpha} \frac{\partial D(T)}{\partial y} \Big|_{u=0} \\ &= 2 \frac{\partial D(T)}{\partial y} \Big|_{y=\beta}, \end{aligned}$$

$$D(T)|_{u=0} = D(T)|_{y=\beta},$$

$$\begin{aligned} Q'(0) &= \int_{-\infty}^{0} \{ \tfrac{3}{2} D(T,0,0)^{-\frac{1}{2}} D(T,0,1) D(T,1,0) - 2 D(T,0,0)^{\frac{3}{2}} D(T,1,1) \} dv, \\ &= \int_{-\infty}^{0} \left\{ -\frac{(T-2)}{2} x^{-\frac{1}{2}(T+3)} - \frac{(T+2)(T-3)}{2} x^{-\frac{1}{2}(T+5)} + \frac{T^2+2T-12}{2} x^{-\frac{1}{2}(T-7)} \right\} dv, \\ &= \frac{12}{(T+1)(T+3)(T+5)}. \end{aligned}$$

In a similar fashion we obtain the remaining coefficients of the Maclaurin series.

$$
\begin{aligned}
Q(0) &= \frac{T^2-2T+3}{(T-1)(T+1)} = 1-\frac{2}{T}+\frac{4}{T^2}-\frac{2}{T^3}+O(T^{-4}),\\
Q'(0) &= \frac{12}{(T+1)(T+3)(T+5)} = \frac{12}{T^3}+O(T^{-4}),\\
Q''(0) &= \frac{36(T+8)}{(T+3)(T+5)(T+7)(T+9)} = \frac{36}{T^3}+O(T^{-4}),\\
E(\hat{\alpha}) &= \left(1-\frac{2}{T}+\frac{4}{T^2}-\frac{2}{T^3}\right)\alpha+\frac{12\alpha^3}{T^3}+\frac{18\alpha^5}{T^3}+\ldots, \qquad (9)\\
R(0) &= \frac{T^2-4T+7}{(T-3)(T-1)(T+1)} = \frac{1}{T}-\frac{1}{T^2}+\frac{5}{T^3}+O(T^{-4}),\\
R'(0) &= \frac{T^4+13T^3-4T^2+21T+63}{(T-1)(T+1)(T+3)(T+5)} = 1-\frac{5}{T}+\frac{22}{T^2}+\frac{77}{T^3}+O(T^{-4}),\\
R''(0) &= \frac{18(3T^2+48T+149)}{(T+1)(T+3)(T+5)(T+7)(T+9)} = \frac{54}{T^3}+O(T^{-4}),\\
E(\hat{\alpha}^2) &= \left(\frac{1}{T}-\frac{1}{T^2}+\frac{5}{T^3}\right)+\left(1-\frac{5}{T}+\frac{22}{T^2}-\frac{77}{T^3}\right)\alpha^2+\frac{27}{T^3}\alpha^4+\ldots.
\end{aligned}
$$

Model B

For model B (4) becomes

$$
\begin{aligned}
E^*(\exp\{Uu+Vv\}) = m^*(u,v) &= \int \exp(Uu+Vv) f^*(x)\,dx\\
&= (2\pi)^{-\frac{1}{2}T}(1-\alpha^2)^{\frac{1}{2}}\int \exp(-\tfrac{1}{2}xD^*x')\,dx\\
&= (1-\alpha^2)^{\frac{1}{2}}\,|D^*|^{-\frac{1}{2}}.
\end{aligned}
$$

Where D^* is the $T\times T$ matrix whose determinant is

$$
D^*(T) = \begin{vmatrix}
p-\beta & q & 0 & \ldots & \ldots & \ldots\\
q & p & q & & & \\
0 & q & p & & & \\
\ldots & \ldots & \ldots & & & \\
 & & & q & p & q\\
 & & & 0 & q & 1
\end{vmatrix}.
$$

$$
p = 1+\alpha^2-2v, \quad q = (\alpha+u), \quad \beta = \alpha^2.
$$

Expanding $D^*(T)$ by the elements of the first row we have

$$
D^*(T) = (p-\beta)\,D(T-1)-D(T-2), \qquad (10)
$$

where $D(T)$ is as defined in (5). Combining (10) and (7) we find that

$$
D^*(T) = D(T)-\beta D(T-1).
$$

Defining

$$
D^*(T,i,j) = \frac{\partial}{\partial\beta^j}\left(\left.\frac{\partial D(T)}{\partial y}\right|_{y=\beta}\right)\Bigg|_{\beta=0}
$$

we have

$$D^*(T,0,0) = x^{T-1},$$

$$D^*(T,0,1) = x^{T-3}\{(T-3)x-(T-2)\},$$

$$D^*(T,0,2) = (T-3)(T-4)x^{T-5}(x-1)^2,$$

$$D^*(T,1,0) = -x^{T-3}\{x+(T-2)\},$$

$$D^*(T,1,1) = -(T-3)x^{T-5}\{x^2+(T-5)x-(T-4)\},$$

$$D^*(T,1,2) = -(T-4)x^{T-7}\{(T-3)x^3+(T-4)(T-7)x^2-(2T-11)(T-5)x + (T-5)(T-6)\},$$

$$D^*(T,2,0) = (T-3)x^{T-5}\{2x+(T-4)\},$$

$$D^*(T,2,1) = (T-4)x^{T-7}\{2(T-4)x^2+(T-5)(T-7)x-(T-5)(T-6)\}.$$

The values of $D^*(T,i,j)$ for $i = 0, 1$ agree with those obtained by Hurwicz (1950, p. 381).

Proceeding as with model A, we expand the respective integrands in their Maclaurin series, integrate and multiply by $(1-\alpha^2)^{\frac{1}{2}}$ to obtain

$$\frac{E^*(\hat{\alpha})}{\alpha} = Q^*(0)+Q^{*\prime}(0)\beta+Q^{*\prime\prime}(0)\frac{\beta^2}{2}+\ldots,$$

$$E^*(\hat{\alpha}^2) = R^*(0)+R^{*\prime}(0)\beta+R^{*\prime\prime}(0)\frac{\beta^2}{2}+\ldots,$$

$$Q^*(0) = \frac{T^2-2T+3}{(T-1)(T+1)} = 1-\frac{2}{T}+\frac{4}{T^4}-\frac{2}{T^3}+O(T^{-4}),$$

$$Q^{*\prime}(0) = \frac{2(T^2+8T-21)}{(T-1)(T+1)(T+3)(T+5)} = \frac{2}{T^2}+\frac{0}{T^3}+O(T^{-4}),$$

$$Q^{*\prime\prime}(0) = 4\frac{T^4+24T^3+98T^2-264T-99}{(T-1)(T+1)(T+3)(T+5)(T+7)(T+9)} = \frac{4}{T^2}+\frac{0}{T^3}+O(T^{-4}),$$

$$E^*(\hat{\alpha}) = \left(1-\frac{2}{T}+\frac{4}{T^2}-\frac{2}{T^3}\right)\alpha+\frac{2\alpha^3}{T^2}+\frac{2\alpha^5}{T^2}+\ldots. \tag{11}$$

$Q^*(0)$ and $Q^{*\prime\prime}(0)$ agree with the values obtained by Hurwicz (1950, p. 381, eq. (4·8) and (4·10)). There apparently is a typographical error in Hurwicz's value for $Q^{*\prime}(0)$ (1950, eq. (4·9)).

The values associated with $E^*(\hat{\alpha}^2)$ are

$$R^*(0) = \frac{T^2-4T+7}{(T-3)(T-1)(T+1)} = \frac{1}{T}-\frac{1}{T^2}+\frac{5}{T^3}+O(T^{-4}),$$

$$R^{*\prime}(0) = \frac{1}{2}\left\{\frac{-1}{(T-3)(T-1)}+\frac{4T+13}{(T-1)(T+1)}+\frac{4T-1}{(T+1)(T+3)}+\frac{-2T-1}{(T+3)(T+5)}+2\right\}$$

$$= 1-\frac{5}{T}+\frac{21}{T^2}-\frac{73}{T^3}+O(T^{-4}),$$

$$R^{*\prime\prime}(0) = \frac{1}{4}\left\{\frac{-1}{(T-3)(T+1)}+\frac{-15}{(T-1)(T+1)}+\frac{60}{(T+1)(T+3)}-\frac{27}{(T+5)(T+7)}+\frac{15}{(T+7)(T+9)}\right\}$$

$$= \frac{8}{T^2}-\frac{40}{T^3}+O(T^{-4}),$$

$$E^*(\hat{\alpha}^2) = \left(\frac{1}{T}-\frac{1}{T^2}+\frac{5}{T^3}\right)+\left(1-\frac{5}{T}+\frac{21}{T^2}-\frac{73}{T^3}\right)\alpha^2+\left(\frac{4}{T^2}-\frac{20}{T^3}\right)\alpha^4+\ldots.$$

VARIANCES

From the series expansion given above we may also compute the variances and the quadratic risks for the two models. We find for model A

$$E(\hat{\alpha}-\alpha)^2 = E(\hat{\alpha}^2) - 2\beta E(\hat{\alpha})/\alpha + \beta$$
$$= \frac{T^2-4T+7}{(T-3)(T-1)(T+1)} - \frac{T^3-6T^2+25T-42}{(T-1)(T+1)(T+3)(T+5)}\beta$$
$$+ \frac{3(T+19)(T-3)}{(T+1)(T+3)(T+5)(T+7)(T+9)}\beta^2 + \ldots,$$

$$E(\hat{\alpha}-\alpha)^2 = \left(\frac{1}{T} - \frac{1}{T^2} + \frac{5}{T^3}\right) - \left(\frac{1}{T} + \frac{14}{T^2} - \frac{73}{T^3}\right)\beta + \frac{3}{T^3}\beta^2 + \ldots,$$

$$\sigma^2 = E(\hat{\alpha}^2) - E(\hat{\alpha})^2 = \left(\frac{1}{T} - \frac{1}{T^2} + \frac{5}{T^3}\right) - \left(\frac{1}{T} - \frac{10}{T^2} + \frac{57}{T^3}\right)\beta + \frac{3\beta^2}{T^3} + \ldots.$$

For model B we obtain

$$E^*(\hat{\alpha}-\alpha) = \frac{T^2-4T+7}{(T-3)(T-1)(T+1)}$$
$$-\frac{1}{2}\left(\frac{+1}{(T-3)(T-1)} - \frac{1}{(T-1)(T+1)} - \frac{11}{(T+1)(T+3)} + \frac{2T+1}{(T+3)(T+3)}\right)\beta$$
$$+\left(-\frac{1}{(T-3)(T-1)(T+1)} + \frac{36}{(T+1)(T+3)(T+5)} - \frac{12}{(T+3)(T+5)(T+7)}\right.$$
$$\left. + \frac{15}{(T+3)(T+7)(T+9)}\right)\frac{\beta^2}{2} + \ldots,$$

$$E^*(\hat{\alpha}-\alpha)^2 = \left(\frac{1}{T} - \frac{1}{T^2} + \frac{5}{T^3}\right) - \left(\frac{1}{T} - \frac{13}{T^2} + \frac{69}{T^3}\right)\beta - \frac{20}{T^3}\beta^2 + \ldots,$$

$$\sigma^{*2} = E^*(\hat{\alpha} - E^*(\hat{\alpha}))^2 = \left(\frac{1}{T} - \frac{1}{T^2} + \frac{5}{T^3}\right) - \left(\frac{1}{T} - \frac{9}{T^2} + \frac{53}{T^3}\right)\beta - \frac{12}{T^3}\beta^2 + \ldots.$$

Thus we have, up to terms of order $O(T^{-2})$

$$\sigma^2 = \sigma^{*2} = E(\hat{\alpha}-\alpha)^2 = E^*(\hat{\alpha}-\alpha)^2 = \frac{1-\alpha^2}{T}. \tag{12}$$

Model A, $x_0 \neq 0$

For model A it is also of interest to examine the moments of $\hat{\alpha}$ when $x_0 = c \neq 0$. In this case the joint moment generating function corresponding to (6) is

$$m(u,v) = D(T)^{-\frac{1}{2}} \exp\left(\frac{c^2}{2}\left\{1 - \frac{D(T+1)}{D(T)}\right\}\right).$$

The first term of the Maclaurin series expansion of $E(\hat{\alpha})/\alpha = Q_c(\beta)$ is

$$Q_c(0) = \int_{-\infty}^{0} \frac{1}{\alpha}\frac{\partial m}{\partial u}\bigg|_{u=0,\,\alpha=0} dv$$
$$= \frac{e^{\frac{1}{2}c^2}}{2}\int_1^\infty e^{-\frac{1}{2}(c^2x)}\left\{x^{-\frac{1}{2}(T+1)} + (T-2+c^2)\,x^{-\frac{1}{2}(T+3)}\right\} dx.$$

Setting $a = \frac{1}{2}(T+1), y = \frac{1}{2}c^2$ and integrating by parts we have

$$Q_c(0) = \tfrac{1}{2}e^y\left\{y^{a-1}\Gamma(1-a,y) - \frac{y^a}{a}(2a+2y-3)\,\Gamma(1-a,y) + (2a+2y-3)\frac{e^{-y}}{a}\right\},$$

where $\Gamma(1-a,y)$ is the incomplete gamma function

$$\Gamma(1-a,y) = \int_y^\infty e^{-u}u^{-a}\,du.$$

From Tricomi's asymptotic expansion for $\Gamma(1-a,y)$ (Erdelyi (1953), p. 140, eq. 9·5 (4)) we have

$$\Gamma(1-a,y) = \frac{e^{-y}y^{1-a}}{y+a}\left\{1+\frac{a}{(y+a)^2}+\frac{2a}{(y+a)^3}+O(\alpha^2(y+a)^{-4})\right\},$$

$$Q_c(0) = \frac{1}{2}\left(\frac{2a+2y-2}{y+a}\right)+O\left(\frac{T^2}{(T+c^2)^4}\right) = 1-\frac{2}{T+1+c^2}+O\left(\frac{T^2}{(T+c^2)^4}\right),$$

$$E(\hat{\alpha}) = \left(1-\frac{2}{T+1+c^2}\right)\alpha+\ldots.$$

We note that for $c = 0$ this agrees with (9). We also note that for large values of $|c| = |x_0|$ we have

$$\frac{E(\hat{\alpha})}{\alpha} \cong 1.$$

This supports a conjecture of Hurwicz (1950, p. 373).

The calculations required for the remaining terms of $E(\hat{\alpha})/\alpha$ and those for $E(\hat{\alpha}^2)$ are too laborious to be attempted at this time.

COMPARISON WITH OTHER RESULTS

Defining the serial correlation coefficient of order k as

$$\alpha_k = \frac{\sum_{t=1}^{n-k} x_t x_{t+k}}{\sum_{t=1}^{n-k} x_t^2},$$

Marriott & Pope (1954) have obtained

$$E^*(\alpha_k) = \alpha^k - \frac{2k\alpha^k}{T}+O(T^{-2}).$$

For $k = 1$ this reduces to $\quad E^*(\hat{\alpha}) = \alpha\left(1-\frac{2}{T}\right)+O(T^{-2})$

which agrees with (11). We note that (9) also gives

$$E(\hat{\alpha}) = \alpha\left(1-\frac{T}{2}\right)+O(T^{-2}).$$

Bartlett has derived the asymptotic variance of α_k as

$$\operatorname{var}(\alpha_k) = \frac{1}{T}\frac{(1+\alpha^2)(1-\alpha^{2k})}{1-\alpha^2} - 2k\alpha^{2k}+O\left(\frac{1}{T^2}\right),$$

$$\operatorname{var}(\alpha_1) = \frac{1-\alpha^2}{T}+O\left(\frac{1}{T^2}\right).$$

This is equivalent to (12).

Several authors (Jenkins, 1956; Kendall, 1957; White, 1957*a*, *b*) have considered Leipnik's distribution (Leipnik, 1947) as an approximation to that of $\hat{\alpha}$.

For Leipnik's distribution we have

$$\left.\begin{aligned} E(\hat{\alpha}) &= \frac{T}{T+2}\alpha \cong \left(1-\frac{2}{T}+\frac{4}{T^2}-\frac{8}{T^3}\right)\alpha, \\ E(\hat{\alpha}^2) &\cong \left(\frac{1}{T}-\frac{2}{T^2}+\frac{4}{T^3}\right)+\left(1-\frac{5}{T}+\frac{22}{T^2}-\frac{92}{T^3}\right)\beta, \\ E(\hat{\alpha}-\alpha)^2 &\cong \left(\frac{1}{T}-\frac{2}{T^2}+\frac{4}{T^3}\right)-\left(\frac{1}{T}-\frac{10}{T^2}+\frac{60}{T^3}\right)\beta. \end{aligned}\right\} \tag{13}$$

Comparing these results with those obtained for model A we note that up to terms of order $O(T^{-3})$ the means are the same, while second moments and variances differ only by $1/T^2+O(T^{-3})$. The agreement with model B is only to terms of order $O(T^{-2})$.

These results support the use of Leipnik's distribution as an approximation for either model A or model B for large T (say $T > 20$). This implies that we may also use approximations derived from Leipnik's distribution such as Quenouille's t-test (Quenouille, 1949; White, 1957*b*)

$$t = (\hat{\alpha}-\alpha)(T+1)^{\frac{1}{2}}(1-\hat{\alpha}^2)^{-\frac{1}{2}} \quad (df = T+1)$$

and Jenkins' angular transformation (Jenkins, 1954)

$$z = \sin^{-1}\hat{\alpha} \sim N(\sin^{-1}\alpha, 1/T).$$

Final remarks

Only the first three terms of the Maclaurin series for the first two moments of $\hat{\alpha}$ have been obtained here. It would be of interest to have available the series for the third and fourth moments. However, the time required to compute the jth item in the expansion of α^k is roughly proportional to 2^{j+k}.

It would also be of interest to examine similar expansions when the mean of the process is unknown and must be estimated. In this case an appropriate estimator might be

$$\bar{\alpha} = \frac{\Sigma(x_t-\bar{x}_0)(x_{t-1}-\bar{x}_1)}{\Sigma(x_{t-1}-\bar{x}_1)^2}, \quad \text{where} \quad \bar{x}_i = \frac{\Sigma x_{t-i}}{T}.$$

We then have (Kendall, 1954; Marriott & Pope, 1954)

$$E^*(\bar{\alpha}) = -\frac{1}{T}+\left(1-\frac{3}{T}\right)\alpha+O(T^{-2})$$

which differs considerably from the expected value of $\hat{\alpha}$.

A third model for serial correlation is the so-called 'circular' case. Here the estimator is defined as

$$\hat{\alpha}_c = \frac{\Sigma x_t x_{t-1}}{\Sigma x_t^2} \quad (x_0 = x_T).$$

Leipnik's distribution was derived as an approximation to the distribution of $\hat{\alpha}_c$ and thus it is reasonable to assume the series expansions for the moments of $\hat{\alpha}_c$ will be similar to those of (13) and hence similar to those for $\hat{\alpha}$.

On the other hand Kendall (1954) has shown that when the mean of x_t is unknown

$$E(\bar{\alpha}_c) = -\frac{1}{T} + \left(1 - \frac{4}{T}\right)\alpha + O\left(\frac{1}{T^2}\right).$$

Comparing this with $E(\bar{\alpha})$ we see that the two expectations differ by $1/T$.

We might also consider the so-called 'explosive' case of model A, i.e. $|\alpha| \geqslant 1$. In this case it has been shown that $\hat{\alpha}$ has an asymptotic Cauchy distribution (White, 1958, 1959), and hence has no asymptotic moments. It seems likely that the formal series expansions given above have a radius of convergence of 1 and hence are only valid for $|\alpha| < 1$.

REFERENCES

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of auto-correlated time series. *J. R. Statist. Soc.* Suppl. **8**, 27.

ERDELYI, A. (Ed.) (1953). *Higher Transcendental Functions*, **2**. New York: McGraw-Hill Inc.

HURWICZ, L. (1950). Least-squares bias in time series. *Statistical Inference in Dynamic Economic Models*, pp. 265–84 (editor T. C. Koopmans). New York: John Wiley and Sons Inc.

JENKINS, G. M. (1954). An angular transformation for the serial correlation coefficient. *Biometrika*, **41**, 261–5.

JENKINS, G. M. (1956). Tests of hypotheses in the linear autoregressive model. II. *Biometrika*, **43**, 186–99.

KENDALL, M. G. (1954). Note on the bias in the estimation of autocorrelation. *Biometrika*, **41**, 403–4.

KENDALL, M. G. (1957). The moments of the Leipnik distribution. *Biometrika*, **44**, 270–2.

KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variables. *Ann. Math. Statist.* **13**, 14–23.

LEIPNIK, R. B. (1947). Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Statist.* **18**, 80–7.

MARRIOTT, F. H. C. & POPE, J. A. (1954). Bias in the estimation of autocorrelation. *Biometrika*, **41**, 390–6.

QUENOUILLE, M. H. (1949). Approximate tests of correlation in time-series 3. *Proc. Camb. Phil. Soc.* **45**, 483–4.

WHITE, J. S. (1957*a*). Approximate moments for the serial correlation coefficient. *Ann. Math. Statist.* **28**, 798–802.

WHITE, J. S. (1957*b*). A t-test for the serial correlation coefficient. *Ann. Math. Statist.* **27**, 1046–8.

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188–97.

WHITE, J. S. (1959). The limiting distribution of the serial correlation coefficient in the explosive case, II. *Ann. Math. Statist.* **30**, 831–4.

WILLIAMS, J. D. (1941). Moments of the ratio of the mean square successive difference in samples from a normal universe. *Ann. Math. Statist.* **12**, 239–41.

Biometrika (1961), **48**, 1 *and* 2, *p.* 95
Printed in Great Britain

Significance tests for paired-comparison experiments†

BY T. H. STARKS AND H. A. DAVID
E. I. du Pont de Nemours and Company, Inc. and Virginia Polytechnic Institute

1. INTRODUCTION

The method of paired comparisons was introduced by Thurstone (1927) as a way of assessing the relative strengths of several 'stimuli' when no meaningful absolute measurements are possible. An important situation to which the method may be applied is that of preference testing, in which the aim is to arrange on some sort of scale the preferences expressed by a panel of judges for a number of treatments subjectively compared two at a time. Although many papers dealing with the analysis of paired-comparison experiments have appeared since 1927 little has been written about tests of significance involving individual treatments or subgroups of treatments. One exception is a paper by Scheffé (1952) who used Tukey's test based on allowances to separate treatments into significantly different groups, but required a scoring of the individual comparison decisions and certain normality assumptions.

For simple (i.e. unrepeated) paired-comparison experiments David (1959) gave significance tests, resting on a minimum of assumptions, for a particular treatment, for the difference between two particular treatments, and for the treatment most often preferred. In the present paper these tests are extended to repeated experiments. Also three multiple comparison procedures for separating significantly different treatment scores are introduced. A numerical example illustrates the use of each of the various methods.

2. NOTATION AND ASSUMPTIONS

We shall consider balanced paired-comparison experiments with t treatments and n repetitions. The score of treatment i ($i = 1, \ldots, t$), i.e. the number of times treatment i is preferred in the experiment, is denoted by a_i. The symbol π_{ij} represents the probability that treatment i is preferred to treatment j in a paired comparison. Ties are not permitted so that

$$\pi_{ji} = 1 - \pi_{ij} \quad (i \neq j, j = 1, \ldots, t). \tag{1}$$

It is assumed that the π_{ij} remain constant throughout the experiment. Thurstone (1927) and Bradley & Terry (1952) propose two models which allow the π_{ij} to be expressed in terms of $t-1$ independent parameters. These are examples of models sometimes called 'linear', since estimates of the treatment effects may usefully be represented by points on a line. The tests developed in the sequel are not tied to a particular linear model, but provide exact or approximate tests under *any* linear model. Indeed, these tests may still be used when no linear model is appropriate, a situation which arises, for example, when $\pi_{ij} > \frac{1}{2}$, $\pi_{jk} > \frac{1}{2}$ but $\pi_{ik} < \frac{1}{2}$.

We shall represent the average preference probability of treatment u ($u = 1, \ldots, t$) by

$$\pi_{u.} = \sum_{j=1}^{t}{}' \pi_{uj}/(t-1), \tag{2}$$

where Σ' indicates summation over all treatments except u.

† This work was sponsored in part by the Office of Ordnance Research, U.S. Army.

The significance test methods in this paper are developed under the null hypothesis H_0' (say) that

$$\pi_{ij} = \tfrac{1}{2} \quad \text{for all } i, j. \tag{3}$$

The tests developed under H_0' are shown to be valid, but occasionally conservative (i.e. the true probability of rejection of the null hypothesis when it is true is less than the stated test significance level), when testing certain more general null hypotheses for which H_0' is a special case.

3. TEST OF ONE PARTICULAR TREATMENT

Although the principal reason for the paired-comparison experiment may be to obtain a ranking of the treatments, the experimenter may be especially interested in whether a particular treatment u is better than the average; that is, he wishes to test

$$H_0\colon \pi_{u.} = \tfrac{1}{2} \tag{4}$$

against

$$H_a\colon \pi_{u.} > \tfrac{1}{2}. \tag{5}$$

Under

$$H_0'\colon \pi_{ij} = \tfrac{1}{2} \quad \text{for all } i, j, \tag{6}$$

the score a_u of treatment u is a binomial variate with parameters $[\frac{1}{2}, n(t-1)]$, and the usual binomial test can be applied to test H_0' versus H_a. If

$$\pi_{u.} = \tfrac{1}{2}, \tag{7}$$

but not all the π_{ui}'s are equal, a_u is a generalized binomial variate (Cramér, 1946, § 16·6) with the same expected value as under H_0', but with a variance that is less than it was under H_0' by the amount

$$n \sum_{i=1}^{t}{}' (\pi_{ui} - \tfrac{1}{2})^2. \tag{8}$$

It can be shown (cf. David, 1960) that the binomial test of H_0 versus H_a will, if it is not exact, be a conservative test procedure.

The binomial property, of course, also can be used to test the alternative hypotheses

$$H_{a2}\colon \pi_{u.} < \tfrac{1}{2} \quad \text{and} \quad H_{a3}\colon \pi_{u.} \neq \tfrac{1}{2}.$$

4. TEST OF EQUALITY OF TWO PRE-ASSIGNED TREATMENTS

Consider the case in which interest is expressed, prior to the performance of the t-treatment experiment, concerning the existence of a distinguishable difference between treatments u and v. After the experiment, we may test

$$\left.\begin{aligned} &H_0\colon \pi_{u.} = \pi_{v.} \\ \text{against} \quad &H_a\colon \pi_{u.} \neq \pi_{v.} \end{aligned}\right\} \tag{9}$$

The test procedure is again developed under

$$H_0'\colon \pi_{ij} = \tfrac{1}{2} \quad \text{for all } (i, j).$$

For any positive integer m, let

$$P_{uvm} = \Pr(|a_u - a_v| \geqslant m | H_0') = 2\Pr(a_u - a_v \geqslant m | H_0')$$

$$= 2 \sum_{d=m}^{n(t-1)} \sum_{p=0}^{n} [\Pr(\text{In all comparisons between treatments } u \text{ and } v,\ u \text{ is preferred } k = (2p-n) \text{ more times than is } v) \times \Pr(\text{treatment } u \text{ is preferred } (d-k) \text{ more times than is } v \text{ in comparisons with the other } (t-2) \text{ treatments})].$$

$$P_{uvm} = 2 \sum_{d=m}^{n(t-1)} \sum_{p=0}^{n} \binom{n}{p} 2^{-n} \sum_{q=d-2p+n}^{n(t-2)} \binom{n(t-2)}{q} \binom{n(t-2)}{q-d+2p-n} 2^{-2n(t-2)}$$

$$= 2^{3n-2nt+1} \sum_{d=m}^{n(t-1)} \sum_{p=0}^{n} \binom{n}{p} \sum_{q=d-2p+n}^{n(t-2)} \binom{n(t-2)}{q} \binom{n(t-2)}{q-d+2p-n}. \tag{10}$$

The equality of the coefficients of $x^{-d+2p-n}$ in the expansion of the two sides of the identity

$$(1+x)^{n(t-2)}(1+x^{-1})^{n(t-2)} \equiv (1+x)^{2n(t-2)} x^{-n(t-2)}, \tag{11}$$

gives

$$\sum_{q=d-2p+n}^{n(t-2)} \binom{n(t-2)}{q} \binom{n(t-2)}{q-d+2p-n} = \binom{2n(t-2)}{n(t-3)-d+2p}. \tag{12}$$

Substituting (12) into (10), we obtain

$$P_{uvm} = 2^{3n-2n+1} \sum_{d=m}^{n(t-1)} \sum_{p=0}^{n} \binom{n}{p} \binom{2n(t-2)}{n(t-3)-d+2p}. \tag{13}$$

Hence to test H_0' against H_a, we follow the

Test procedure

(1) Choose the desired significance level α.

(2) Find m_c the smallest integer value of m for which P_{uvm} does not exceed α.

(3) Accept H_a if $|a_u - a_v|$ is not less than m_c.

To test H_0' against the one-sided alternative hypothesis

$$H_{a2}\colon\ \pi_{u.} > \pi_{v.}$$

use the above test procedure with $\frac{1}{2}P_{uvm}$ in place of P_{uvm} in Step 2 and remove the absolute value signs in Step 3. The critical values of m for one- and two-sided tests when $\alpha = 0{\cdot}01$ or $0{\cdot}05$ are given in Table 1.

As was pointed out by David (1959) and as observed in Table 1, the distribution of the difference

$$m = a_u - a_v$$

tends rapidly toward the normal $(0, \frac{1}{2}nt)$ distribution as either n or t, or both, become large. The convergence follows from the fact that the characteristic function of $m/\sqrt{(\frac{1}{2}nt)}$ is

$$\Phi_{m/\sqrt{(\frac{1}{2}nt)}}(\omega) = [\cos \tfrac{1}{2}w\sqrt{(2/nt)}]^{2n(t-2)} [\cos w\sqrt{(2/nt)}]^{n}, \tag{14}$$

which tends to $e^{-\frac{1}{2}w^2}$ as n or t become large.

For the same reasons as mentioned in the preceding section, this test becomes conservative when H_0 is true, but differs from H_0'.

Table 1. *Critical values for the difference between scores of two pre-assigned treatments*

Experiment size		$\alpha = 0{\cdot}01$		$\alpha = 0{\cdot}05$	
n	t	One-sided test m_e'	Two-sided test m_e	One-sided test m_e'	Two-sided test m_e
1	$\leqslant 4$	No significant values		No significant values	
1	5	4	None possible	4	4
1	6	5	5	4	4
1	7	5	5	4	5
1	8	5	6	4	5
1	9	6	6	4	5
1	10	6	7	5	5
1	11	6	7	5	6
1	12	7	7	5	6
1	13	7	7	5	6
1	14	7	8	5	6
1	15	7	8	5	6
1	16	7	8	6	6
2	3	No significant values		4	4
2	4	5	6	4	5
2	5	6	6	5	5
3	3	6	6	4	5
3	4	6†	7	5	6
4	3	6‡	7	5	6
4	4	7	8	6	6
All larger values of n or t		m_e' = smallest integer $\geqslant 2{\cdot}33\sigma + 0{\cdot}5$	m_e = smallest integer $\geqslant 2{\cdot}56\sigma + 0{\cdot}5$	m_e' = smallest integer $\geqslant 1{\cdot}64\sigma + 0{\cdot}5$	m_e = smallest integer $\geqslant 1{\cdot}96\sigma + 0{\cdot}5$

$[\sigma = (\frac{1}{2}nt)^{\frac{1}{2}}]$

† $P_{uv6} = 0{\cdot}0103$. ‡ $P_{uv6} = 0{\cdot}01001$.

5. A TEST OF THE HIGHEST SCORE

After running a paired-comparison experiment, the experimenter often wishes to know whether the treatment with the highest score is distinguishably better than the average of the t treatments. If the treatment with the largest score is labelled as treatment (1), he wishes to test the null hypothesis

$$H_0\colon \pi_{(1).} = \tfrac{1}{2}$$

against the alternative hypothesis $$H_a\colon \pi_{(1).} > \tfrac{1}{2}. \qquad (15)$$

Again we develop the test under $H_0'\colon \pi_{ij} = \frac{1}{2}$ for all (i, j). To perform a test of H_0' against H_a, one needs information concerning at least the critical part of the distribution of the largest score, $a_{(1)}$, under H_0'.

If we let A_i be the event $a_i \geqslant M$ $[0 \leqslant M \leqslant n(t-1)]$, then the elementary law of probability concerning the sum of events, plus the equi-probability of the events A_i under H'_0, yields

$$\Pr(a_{(1)} \geqslant M) = \Pr\left(\sum_{i=1}^{t} A_i\right) = \sum_{j=1}^{t} (-1)^{j-1}\binom{t}{j} \Pr(A_1 A_2 \dots A_j). \tag{16}$$

Aside from the trivial case $M = 0$, there exists one other set of values of M, namely, $M > n(t-1) - \frac{1}{2}n$, for which it is easy to evalute $\Pr(a_{(1)} \geqslant M)$. When $M > n(t-1) - \frac{1}{2}n$,

$$\Pr(a_{(1)} \geqslant M | H'_0) = \binom{t}{1} \Pr(A_i) = t.2^{-n(t-1)} \sum_{k=M}^{n(t-1)} \binom{n(t-1)}{k}, \tag{17}$$

since no two treatment scores can simultaneously exceed $[n(t-1) - \frac{1}{2}n]$.

When

$$0 < M \leqslant n(t-1) - \tfrac{1}{2}n, \tag{18}$$

it is difficult to determine the joint probabilities on the right-hand side of (16) since the scores are correlated binomial variates. However, we can use a simple approximation. For the values of M such that $\Pr(A_i) < 1/t$, one can apply the Bonferroni inequality to obtain

$$t.\Pr(A_i) - \binom{t}{2}.\Pr(A_i A_j) \leqslant \Pr(a_{(1)} \geqslant M) \leqslant t.\Pr(A_i). \tag{19}$$

Since the sum of the treatment scores is a constant, we have

$$\Pr(A_i | A_j) \leqslant \Pr(A_i), \tag{20}$$

and therefore,

$$\Pr(A_i A_j) \leqslant [\Pr(A_i)]^2. \tag{21}$$

It follows that

$$t.\Pr(A_i) - \binom{t}{2}[\Pr(A_i)]^2 \leqslant \Pr(a_{(1)} \geqslant M) \leqslant t.\Pr(A_i). \tag{22}$$

Since

$$\Pr(A_i) = 2^{n(t-1)} \sum_{k=m}^{n(t-1)} \binom{n(t-1)}{k} \tag{23}$$

can be calculated directly or found in a table of the binomial probability distributions, we have limits on $\Pr(a_{(1)} \geqslant M | H'_0)$ that are easy to obtain.

To make an approximate α-level significance test of H'_0 against H_a, one should choose as the critical value that positive integer value of M, say M_β, for which

$$t.\Pr(a_i \geqslant M_\beta | H'_0) = \beta \leqslant \alpha < t.\Pr(a_i \geqslant M_\beta - 1 | H'_0). \tag{24}$$

When $\beta = 0{\cdot}05$ and t is large, we find from relations (22) and (24) that

$$0{\cdot}04875 \leqslant \Pr(a_{(1)} \geqslant M_{(0{\cdot}05)} | H'_0) \leqslant 0{\cdot}05. \tag{25}$$

The range between the limits decreases as t or β, or both, decrease. Hence the true significance level of the above approximate test of H'_0 versus H_a is known to be between quite narrow bounds.

If more accuracy is necessary in testing H'_0 against H_a, one may consider the general form of the joint density function of s $(s \leqslant t)$ scores

$$f(a_i, a_j, \dots, a_k) = 2^{\frac{1}{2}ns(2t-s-1)} \sum \prod_{p}^{i,j\dots,k} \binom{t-s}{a_p - a'_p}, \tag{26}$$

where a'_p is the score of treatment p in a subexperiment between the s treatments, and the sum is over all outcomes of subexperiments compatible with the final score. By using (26)

to find the joint probabilities on the right-hand side of (16), one obtains the exact value of $\Pr(a_{(1)} \geqslant M|H_0')$. It should be observed that this method will become increasingly difficult as n and t become large.

Again because the a_i change from binomial to generalized binomial variates when we change from H_0' to H_0, we have that the following test of H_0 versus H_a is a conservative test when H_0 is true but H_0' is not.

Test procedure

(1) Choose the desired significance level α.

(2) From a table of the cumulative binomial probability distribution, find the integer M_β such that

$$t.\Pr(a_i \geqslant M_\beta|H_0') = \beta \leqslant \alpha < t.\Pr(a_i \geqslant [M_\beta - 1]|H_0').$$

(3) If the largest score in the experiment is not less than M_β, accept the hypothesis that the treatment corresponding to that score is better than average.

A similar argument would give the approximate significance test of whether the treatment with the smallest score is below the average of the t treatments.

6. Multiple comparison tests of treatment scores

To separate significantly different treatment scores, two methods are introduced. One is a multiple comparison range test analogous to Tukey's test based on allowances (see Federer, 1955, p. 29) for separating significantly different independent normally distributed sample means. The other method is analogous to Fisher's least significant difference method.

6·1. *Multiple comparison range test*

The multiple range method for a paired-comparison experiment is as follows:

(1) Choose a significance level α.

(2) Find a positive integer $R_{\beta(\alpha)}$, say, such that

$$\Pr[a_{(1)} - a_{(l)} \geqslant R_{\beta(\alpha)}] = \beta \leqslant \alpha < \Pr[a_{(1)} - a_{(l)} \geqslant R_{\beta(\alpha)} - 1)], \tag{27}$$

where $a_{(i)}$ denotes the ith largest score and the probabilities are calculated under the hypothesis

$$H_0'\colon\ \pi_{ij} = \tfrac{1}{2} \quad (i \neq j;\ i, j = 1, \ldots, t).$$

(3) Any pairwise difference in scores not less than $R_{\beta(\alpha)}$ is considered significant.

This test is based on the argument that if any two scores differ by more than we could reasonably expect the two extreme scores to differ by under H_0', we should declare the pair significantly different. This test is conservative when the model is not linear (i.e. when $H_0\colon \pi_{i.} = \frac{1}{2}$, for all i, is not equivalent to H_0').

Step 2 of the above test procedure requires a knowledge of the distribution of the range $a_{(1)} - a_{(l)}$, of the t scores under H_0'. There are $t(t-1)$ differences of the form $(a_i - a_j)$; we are interested in the distribution of the maximum of the $t(t-1)$ differences. Equation (16) expresses the probability that the maximum of t scores will not be less than a certain quantity M as the sum of k-fold $(k = 1, \ldots, t)$ joint probabilities. If we let R_{ij} be the event $(a_i - a_j) \geqslant R$, where R is a positive integer, then the substitution of $t(t-1)$ and the R_{ij}'s in the right-hand side of (16) for t and the A_i's, respectively, gives an expression for $\Pr(a_{(1)} - a_{(l)} \geqslant R)$. When

$$R > n(t-1) - \tfrac{1}{2}n, \tag{28}$$

the R_{ij}'s are mutually exclusive events; and, since every R_{ij} is equally likely under H'_0,

$$\Pr(a_{(1)}-a_{(0)} \geqslant R|H'_0,\ R > n(t-1)-\tfrac{1}{2}n) = t(t-1)\Pr(R_{ij}) \tag{29}$$

and by (13)
$$\Pr(R_{ij}) = 2^{3n-2nt}\sum_{d=R}^{n(t-1)}\sum_{p=0}^{n}\binom{n}{p}\binom{2n(t-2)}{n(t-3)-R+2p}. \tag{30}$$

However, to find $\Pr(a_{(1)}-a_{(0)} \geqslant R)$ when

$$0 < R \leqslant n(t-1)-\tfrac{1}{2}n, \tag{31}$$

it is necessary to evaluate joint probabilities of the form

$$\Pr(R_{ij} \ldots R_{gh}). \tag{32}$$

The nature of the R_{ij}'s makes the evaluation of (32) difficult. For this reason, an approximation to the distribution of $(a_{(1)}-a_{(0)})$ is obtained that is reasonably accurate in the critical part of the distribution (i.e. $0{\cdot}005 \leqslant \alpha \leqslant 0{\cdot}075$). If we consider

$$d_i = 2[a_i-\tfrac{1}{2}n(t-1)](nt)^{-\frac{1}{2}} \quad (i = 1, \ldots, t), \tag{33}$$

then we have that under H'_0
$$E(d_i) = 0, \tag{34}$$

$$E(d_i^2) = (t-1)/t, \tag{35}$$

$$E(d_i d_j) = -1/t \quad (i \neq j). \tag{36}$$

Also, since a_i is a binomial variate, d_i is asymptotically normal. Therefore we may expect the distribution of the range (cf. Hartley, 1950)

$$d_{(1)}-d_{(0)} = 2(a_{(1)}-a_{(0)})/\sqrt{(nt)} \tag{37}$$

to be asymptotically the same as that for t independent observations from a normal population with variance
$$\sigma^2(1-\rho) = (t-1)/t+1/t = 1. \tag{38}$$

The probability integral of the range W_t (say) of t independent observations ($t = 2(1)20$) from a normal population with variance one is given in Table 23 in the *Biometrika Tables for Statisticians*, Vol. 1, by Pearson & Hartley (1954). Hence $\Pr(W_t \geqslant 2R/\sqrt{(nt)})$ is an easily obtained approximation to $\Pr(a_{(1)}-a_{(0)} \geqslant R)$ and will improve in accuracy as $n(t-1)$ becomes large.

The range W_t is a continuous variate; whereas the range $(d_{(1)}-d_{(0)})$ is a discrete variate with distinct values differing by not less than $2(nt)^{-\frac{1}{2}}$. This suggests that a continuity correction might improve the approximation. The usual correction would subtract $(nt)^{-\frac{1}{2}}$ from the observed range of the d_i's, $2R(nt)^{-\frac{1}{2}}$. However, it has been empirically found that more accurate results are obtained in the interval of significance levels $(0{\cdot}005, 0{\cdot}075)$ through the subtraction of $\frac{1}{2}(nt)^{-\frac{1}{2}}$ from $2R(nt)^{-\frac{1}{2}}$.

For small experiment sizes, $\Pr(a_{(1)}-a_{(0)} \geqslant R|H'_0)$ can be obtained from the tables, giving the probabilities of all possible outcomes, found in Bradley & Terry (1952), Bradley (1954), and David (1959). In Table 2, the approximation method given above is compared with values obtained from the tables for the largest experiments given in the tables. Due to the asymptotic nature of the distribution of $a_{(1)}-a_{(0)}$, we may expect the approximation to be better for larger experiments.

To summarize, we shall restate the steps in this multiple range test of treatment scores.

Table 2. *Comparison of* $\Pr(a_{(1)}-a_{(0)} \geqslant R)$ *with its approximation* $\Pr(W_t \geqslant (2R-\frac{1}{2})(nt)^{-\frac{1}{2}})$

t	n	R	$\Pr(a_{(1)}-a_{(0)} \geqslant R\|H_0')$†	$\Pr(W_t \geqslant (2R-\frac{1}{2})(nt)^{-\frac{1}{2}})$
3	10	12	0·0062 ± 0·00003	0·0078
3	10	11	·0157	·0152
3	10	10	·0340 ± ·00003	·0317
4	6	11	·0094	·0104
4	6	10	·0258	·0252
4	6	9	0·0616	0·0560
4	7	12	·0085 ± ·00008	·0092
4	7	11	·0217 ± ·00006	·0212
4	7	10	·0493 ± ·00005	·0454
4	8	13	·0075 ± ·00009	·0078
4	8	12	0·0178 ± 0·00008	0·0174
4	8	11	·0389 ± ·00006	·0363
5	5	12	·0071 ± ·00015	·0079
5	5	11	·0199 ± ·00013	·0200
5	5	10	·0494 ± ·00011	·0460
8	1	7	0·0068	0·0169
8	1	6	·0738	·0778

† The values of $\Pr(a_{(1)}-a_{(0)} \geqslant R|H_0')$ were calculated from the tables of Bradley & Terry (1952), Bradley (1954) and David (1959). Because of the construction of the Bradley–Terry and Bradley tables, it was sometimes necessary to subtract probabilities of smaller ranges from the cumulative probability entered in these tables. Each subtraction could cause errors of as much as ±0·00005. The entries in such cases are of the form $\beta \pm \sigma$, where $\sigma = 0{\cdot}00005\ (k/3)$, and k represents the number of subtractions.

Test procedure

(1) Choose the desired significance level α.

(2) Find a positive integer $R_{\beta(\alpha)}$, say, such that

$$\Pr[a_{(1)}-a_{(0)} \geqslant R_{\beta(\alpha)}] = \beta \leqslant \alpha < \Pr[a_{(1)}-a_{(0)} \geqslant R_{\beta(\alpha)}-1], \tag{39}$$

where the probabilities are calculated under hypothesis H_0'.

(*a*) If the experiment is small, use Table 3 to obtain $R_{\beta(\alpha)}$.

(*b*) If the experiment is too large for method (*a*), find the upper $100\alpha\,\%$ point, $W_t(\alpha)$ say, of the W_t distribution (*Biometrika Tables*, Table 22) and solve

$$W_{t(\alpha)} = (2R^*-\tfrac{1}{2})(nt)^{-\frac{1}{2}} \tag{40}$$

for R^*. If R^+, the smallest integer not less than R^*, is larger than $n(t-1)-\frac{1}{2}n$, use equation (27) to obtain β and $R_{\beta(\alpha)}$; otherwise, set

$$R_{\beta(\alpha)} = R^+. \tag{41}$$

(3) Any pairwise difference in scores not less than $R_{\beta(\alpha)}$ will be considered a significant difference.

This method is easily extended to large balanced subjective experiments with more than two treatments per block by using a generalized form of the d_i-statistic in the previous argument. Let the blocks be of size $k(2 \leqslant k \leqslant t)$ and the treatments in each block be scored from 1 to k, then the generalized form of d_i, say d_i^*, is

$$d_i^* = 2\sqrt{3}\,(a_i-\bar{a})/[\lambda(k+1)t]^{\frac{1}{2}}, \tag{42}$$

where

$$\bar{a} = \frac{1}{t}\Sigma a_i \tag{43}$$

and λ is the number of blocks in which any particular pair occurs. The critical value of the test is R^+.

Table 3. *Critical values $R_{\beta(\alpha)}$ for the multiple comparison range test*†

		$\alpha = 0{\cdot}01$		$\alpha = 0{\cdot}05$	
t	n	$R_{\beta(\alpha)}$	β	$R_{\beta(\alpha)}$	β
3	1	None	None	None	None
3	2	None	None	None	None
3	3	6	0·01	None	None
3	4	7	·01	6	0·05
3	5	8	·01	7	·04
3	6	9	0·01	8	0·03
3	7	10	·01	8	·05
3	8	10	·01	9	·03
3	9	11	·01	9	·05
3	10	12	·01	10	·03
4	1	None	None	None	None
4	2	6	0·01	None	None
4	3	8	·005	7	0·03
4	4	9	·01	8	·03
4	5	10	·01	9	·03
4	6	11	0·01	9	0·06
4	7	12	·01	10	·05
4	8	13	·01	11	·04
5	1	None	None	None	None
5	2	7	·0154	6	·08
5	3	9	0·01	8	0·04
5	4	11	·005	9	·05
5	5	12	·01	10	·05
6	1	None	None	5	·0586
7	1	None	None	6	·0205
8	1	7	·0068	6	·0738

† These critical values $R_{\beta(\alpha)}$ have been chosen in such a way as to make β as close to α as possible rather than requiring that $R_{\beta(\alpha)}$ be such that $\beta \leqslant \alpha$ as was done in the defining equation (27).

6·2. *Least significant difference method*

A second and less conservative method for comparing treatment scores is one that is analogous to Fisher's least significant difference method of comparing sample means in the analysis of variance. This method consists of first testing

$$H_0\colon\ \pi_{i.} = \tfrac{1}{2} \quad \text{for all } i,$$

against

$$H_a\colon\ \pi_{i.} \neq \tfrac{1}{2} \quad \text{for some } i.$$

(This is analogous to the F-test in the analysis of variance.) For small experiments, the previously mentioned tables of Bradley & Terry, Bradley, and David may be used to test H_0 against H_a. For any experiments outside the range of these tables, the approximate

χ^2-test suggested by Durbin (1951) has been found to be reasonably accurate (cf. Starks, 1958). In terms of the notation used in this paper, the test consists of comparing

$$D = \sum_{i=1}^{t} d_i^2 = 4\left[\sum_{i=1}^{t} (a_i - \bar{a})^2\right] \Big/ nt \tag{44}$$

with the upper $100\alpha\,\%$ point of the χ^2-distribution with $(t-1)$ degrees of freedom.

If one fails to reject H_0, no significant difference between treatment scores is declared. If H_0 is rejected, we may apply the 'test of equality of two pre-assigned treatments', described in § 4 of this paper, to every pair of treatment scores. This is analogous to the t-test in Fisher's least significant difference method.

Test procedure

(1) Choose the desired significance level α.

(2) (*a*) For small experiments, use the available methods and tables of Bradley & Terry, and David, to test whether a difference between treatments exists.

(*b*) For larger experiments, use Durbin's method; that is, compare

$$D = \sum_{i=1}^{t} d_i^2 = 4\left[\sum_{i=1}^{t} a_i^2 - \tfrac{1}{4}tn^2(t-1)^2\right] \Big/ (nt) \tag{45}$$

with the upper $100\alpha\,\%$ point of the χ^2-distribution with $(t-1)$ degrees of freedom and reject the hypothesis of equivalence of treatments if D is greater than the critical χ^2-value.

(3) If no significant difference between treatments is found in 2, the test is completed. If the test in 2 shows significance, find the critical value m_c for a two-sided test on two pre-assigned treatments described above and declare every pair of treatment scores differing by as much as m_c to be significantly different.

7. A METHOD FOR JUDGING CONTRASTS OF TREATMENT SCORES

An argument analogous to that employed by Scheffé (1953) to test contrasts of means in the analysis of variance can be used to develop a method for judging contrasts of treatment scores after finding the observed value of D in the paired-comparison experiment to be significant. Consider the set of orthogonal contrasts

$$Q_k = \sum_{i=1}^{t} L_{i(k)} d_i \quad (k = 1, \ldots, t-1), \tag{46}$$

where

$$\sum_{i=1}^{t} L_{i(k)} = 0, \tag{47}$$

and d_i is as defined in equation (33). Using relations (34), (35), and (36), we obtain the variance of Q_k to be

$$S_k = \sum_{i=1}^{t} \sum_{j=1}^{t} \rho_{ij} L_{i(k)} L_{j(k)} \sigma_i^2 = \sum_{i=1}^{t} L_{i(k)}^2. \tag{48}$$

From the theory of orthogonal contrasts

$$D = \sum_{i=1}^{t} d_i^2 = \sum_{k=1}^{t-1} Q_k^2 / S_k \tag{49}$$

for every set $\{Q_k\}$ of $(t-1)$ mutually orthogonal contrasts. Since the terms on the right-hand side of (49) are all non-negative, we have

$$D \geqslant Q_k^2 / S_k \quad (k = 1, \ldots, t-1), \tag{50}$$

with equality occurring when $L_{i(k)} = d_i$ $(i = 1, \ldots, t)$. Hence, if D_α is the upper $100\alpha\,\%$ point of the distribution of D under H_0',

$$\Pr(D \leqslant D_\alpha) = \Pr(\text{for all possible contrasts, } Q_k,\ Q_k^2/S_k \leqslant D) = 1-\alpha. \tag{51}$$

Hence, after performing the D-test and rejecting

$$H_0\colon\ \pi_{i.} = \tfrac{1}{2} \quad (\text{for all } i) \tag{52}$$

at the α level of significance, the experimenter may judge all contrasts of treatment scores in which he is interested with the following.

Test procedure

(1) Calculate $$Q^2 = \left(\sum_{i=1}^{t} L_i d_i\right)^2 = 4\left(\sum_{i=1}^{t} L_i a_i\right)^2 \Big/ (nt),$$

where $$\sum_{i=1}^{t} L_i = 0. \tag{53}$$

(2) Calculate $$S = \sum_{i=1}^{t} L_i^2. \tag{54}$$

(3) Calculate SD_α, where D_α is the critical value used in the D-test.

(4) If $Q^2 \geqslant SD_\alpha$, declare the contrast significantly different from zero.

For small experiments, α and D_α are determined exactly from tabled distributions and, for larger experiments, the approximation $D_\alpha = \chi^2_{(t-1),\alpha}$, the upper $100\alpha\,\%$ point of the χ^2 distribution with $(t-1)$ degrees of freedom, is used.

8. Examples

To illustrate the test procedures presented above, they are applied to data from a paired-comparison experiment run on five brands of carbon paper and described by Fleckenstein, Freund & Jackson (1958). The experiment and analysis was carried out according to the method described by Scheffé (1952). Since the test methods in this paper do not allow for scaling or ties, the degree of preference is ignored and merely recorded as a preference and ties were randomly assigned as preferences to one of the two members of each tied pair. The number of times each paper is preferred in the thirty repetitions of the experiment is listed below:

$$a_1 = 66; \quad a_2 = 51; \quad a_3 = 89; \quad a_4 = 24; \quad a_5 = 70.$$

We have arbitrarily decided to run all tests at the 5% significance level.

8·1. *Test of pre-assigned treatment*

Suppose that before the carbon paper experiment, brand 5 had received considerable recommendation because of its low cost and reputed quality. In such a case, the experimenter may be particularly interested in testing whether brand 5 is better than the average of the five brands. The test of $H_0\colon \pi_{5.} = \frac{1}{2}$ against $H_a\colon \pi_{5.} > \frac{1}{2}$ would proceed as follows:

(1) Significance level: 5 %.

(2) Since $n(t-1) = 30(4) = 120$, we may use the normal approximation to find the critical value, a_c, for the score of the pre-assigned treatment

$$a_c = 1{\cdot}64[n(t-1)/4]^{\frac{1}{2}} + \tfrac{1}{2}n(t-1) + \tfrac{1}{2} = 69{\cdot}48.$$

(3) $a_5 = 70$.

(4) $a_5 > a_c$. (The actual significance level of a_5 is 0·0412 under H_0'.) Our conclusion is that brand 5 is better than the average of the five brands.

8·2. *Test of equality of two pre-assigned brands*

If brand 4 were less expensive than brand 2, we might suppose that there would be interest expressed prior to the experiment on whether or not brand 2 is actually superior to brand 4. The one-sided test of H_0: $\pi_{4.} = \pi_{2.}$ against H_a: $\pi_{2.} > \pi_{4.}$ would proceed as follows:

(1) Significance level: 5 %.

(2) $1{\cdot}64(nt/2)^{\frac{1}{2}} + 0{\cdot}5 = 1{\cdot}64\sqrt{75} + 0{\cdot}5 = 14{\cdot}7$, $m_c = 15$.

(3) $a_2 - a_4 = 51 - 24 = 27 > m_c$. We accept the alternative hypothesis H_a that brand 2 is superior to brand 4.

8·3. *Test of the highest score*

It is natural for the experimenter to wonder whether the highest score $a_{(1)}$, which in this example is the score of brand 3, is actually significantly larger than average; that is, if $\pi_{(1).} > \frac{1}{2}$. The test of the null hypothesis against this alternative is as follows:

(1) Significance level: 5 %.

(2) From the binomial tables (Harvard University, 1955), we find for $n = 120$ and $p = \frac{1}{2}$, that

$$\Pr(a_i \geqslant 74 | H_0') = 0{\cdot}033/5,$$

and

$$\Pr(a_i \geqslant 73 | H_0') > 0{\cdot}05/5.$$

Therefore, $M_\beta = 74$ and $\beta = 0{\cdot}033$.

(3) $a_3 = a_{(1)} = 89 > M_\beta$. Our conclusion is that brand 3 is significantly better than average.

8·4. *Multiple comparisons*

To separate the significantly different scores into groups, the experimenter might use the multiple range test, the least significant difference method, or the method of contrasts.

Multiple comparison range test

(1) Significance level: 5 %.

(2) $W_{5,0{\cdot}05} = 3{\cdot}86$ (obtained from Table 22, *Biometrika Tables for Statisticians*).

$$R^* = \tfrac{1}{2}W_{5,0{\cdot}05}\sqrt{(nt)} + \tfrac{1}{4} = 3{\cdot}86\sqrt{(150/4)} + \tfrac{1}{4} = 23{\cdot}887,$$

$$R^+ = 24 < n(t-1) - \tfrac{1}{2}n = 105,$$

$$\beta \doteq \Pr[W_t \geqslant (2R^* - \tfrac{1}{2})(nt)^{-\frac{1}{2}}] = \Pr[W_5 \geqslant 3{\cdot}878] = 0{\cdot}048$$

(from Table 23, *Biometrika Tables for Statisticians*.)

(3)

a_4	a_2	a_1	a_5	a_3
24	51	66	70	89

(4) Any two brands whose scores are not underlined by the same line in the above step may be considered distinguishably different.

Least significant difference method

(1) Significance level: 5 %.

(2)
$$D = 4\{\Sigma a_i^2 - \tfrac{1}{4}tn^2(t-1)^2\}/(nt)$$
$$= 4\{20{,}354 - 18{,}000\}/150 = 62{\cdot}77,$$
$$\chi^2_{4,0{\cdot}05} = 9{\cdot}488 < D.$$

Hence a significant difference exists between the scores.

(3)
$$1{\cdot}96(\tfrac{1}{2}nt)^{\frac{1}{2}} + \tfrac{1}{2} = 1{\cdot}96\sqrt{75} + \tfrac{1}{2} = 17{\cdot}47,$$
$$m_c = 18,$$

a_4	a_2	a_1	a_5	a_3
24	51	66	70	89

Any two brands whose scores are not underlined by the same line may be considered distinguishably different.

Method of contrasts

We have found in Step 2 of the L.S.D. method that a significant difference exists between brands using a D-test with critical value

$$D_{0{\cdot}05} = \chi^2_{4,0{\cdot}05} = 9{\cdot}488.$$

Hence the steps in the contrasts method are as follows:

(1) The separation into groups is accomplished in this case by computing the squares of four contrasts.
$$Q_1^2 = 4(a_3 - a_2)^2/nt = 4(89-51)^2/150 = 38{\cdot}5,$$
$$Q_2^2 = 4(a_3 - a_1)^2/nt = 4(89-66)^2/150 = 14{\cdot}1,$$
$$Q_3^2 = 4(a_5 - a_2)^2/nt = 4(70-51)^2/150 = 9{\cdot}6,$$
$$Q_4^2 = 4(a_2 - a_4)^2/nt = 4(51-24)^2/150 = 19{\cdot}4.$$

(2) For each contrast
$$S = \sum_{i=1}^{2} L_i^2 = 1^2 + 1^2 = 2.$$

(3)
$$SD_\alpha = 2(9{\cdot}488) = 18{\cdot}976.$$

(4)
$$Q_1^2 > Q_4^2 > SD_\alpha > Q_2^2 > Q_3^2.$$

Hence,

a_4	a_2	a_1	a_5	a_3
24	51	66	70	89

Any two brands whose scores are not underlined by the same line may be considered significantly different.

9. Summary

The tests for paired-comparison experiments presented in this paper have dealt with the evaluation of individual treatments, pairs of treatments, the treatment with the highest score, contrasts of treatment scores, and the separation of treatment effects. The methods,

in the main, are based on the binomial distribution and its asymptotic approximations. These methods are developed under the null hypothesis

$$H_0': \pi_{ij} = \tfrac{1}{2} \quad \text{for all } i, j$$

expressing treatment equality. When the actual null hypothesis of the problem (e.g. $H_0: \pi_{u.} = \pi_{v.}$, when testing for a difference between two particular treatments) is true and H_0' is not true, the test based on H_0' is a conservative test of H_0, that is, the probability of exceeding the critical value is less than it was under H_0'.

The three methods developed for the multiple comparison of treatment scores are analogous to methods used for the multiple comparison of sample means in the analysis of variance. The three methods will not, in general, lead to the same interpretation of the paired-comparison data for the same reasons that their analogues in the analysis of variance do not, in general, give the same results. (For a discussion of this subject, the reader is referred to an expository paper by Duncan, 1955.) Also it is shown that one of these methods, the multiple comparison range test, is easily extended for similar analyses of balanced k ($2 \leqslant k \leqslant t$) treatments per block subjective experiments.

The authors wish to express their gratitude to Prof. R. A. Bradley for valuable suggestions and criticisms.

REFERENCES

BRADLEY, R. A. (1954). Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502–37.

BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324–45.

CRÁMER, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

DAVID, H. A. (1959). Tournaments and paired comparisons. *Biometrika*, **46**, 139–49.

DAVID, H. A. (1960). A conservative property of binomial tests. *Ann. Math. Statist.* **31**, 1205–7.

DUNCAN, D. B. (1955). Multiple range and multiple F tests. *Biometrics*, **11**, 1–42.

DURBIN, J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Psychol.* (*Statist. Sect.*), **4**, 85–90.

FEDERER, W. T. (1955). *Experimental Design*. New York: The Macmillan Company.

FLECKENSTEIN, M., FREUND, R. A. & JACKSON, J. E. (1958). A paired comparison test of typewriter carbon papers. *Tappi*, **41**, 128–30.

HARTLEY, H. O. (1950). The use of range in analysis of variance. *Biometrika*, **37**, 271–80.

HARVARD UNIVERSITY COMPUTATION LABORATORY (1955). *Tables of the Cumulative Binomial Probability Distribution*. Cambridge, Mass.: Harvard University Press.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

SCHEFFÉ, H. (1952). An analysis of variance for paired comparisons. *J. Amer. statist. Ass.* **47**, 381–400.

SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87–104.

STARKS, T. H. (1958). Significance tests in experiments involving paired comparisons. Ph.D. dissertation, Virginia Polytechnic Institute Library, Blacksburg, Virginia.

THURSTONE, L. L. (1927). Psychophysical analysis. *Amer. J. Psychol.* **38**, 368–89.

Biometrika (1961), **48**, 1 *and* 2, *p.* 109
Printed in Great Britain

Goodness-of-fit tests on a circle

BY G. S. WATSON†
University of Toronto and Research Triangle Institute

1. INTRODUCTION

To test whether a random sample $X_1, \ldots, X_N$ has been drawn from a population with a specified continuous distribution function, $F(x)$, some measure of the difference between this and the sample distribution function, $F_N(x)$, may be used. Well-known test statistics of this nature are

$$W_N^2 = N \int_{-\infty}^{\infty} \{F_N(x) - F(x)\}^2 \, dF(x) \tag{1}$$

and

$$K_N = \sqrt{N} \sup_{-\infty \leqslant x \leqslant \infty} |F_N(x) - F(x)|. \tag{2}$$

A summary of what is known of these tests and extensions of them is given by Darling (1957). The statistic (1) is associated with the names of Cramér, Von Mises and Smirnov and the latter with those of Kolmogorov and Smirnov. For practical use, both (1) and (2) are available. Neither test has been extended for use when $F(x)$ contains parameters requiring estimation.

In an investigation of the Smirnov statistic, (1), it was observed that while the limiting distribution is very awkward and requires electronic computing for its tabulation, the distribution of

$$U_N^2 = N \int_{-\infty}^{\infty} \left\{F_N(x) - F(x) - \int_{-\infty}^{\infty} [F_N(y) - F(y)] \, dF(y)\right\}^2 dF(x) \tag{3}$$

is very simple and hardly requires tabulation, as will be shown in § 2: This at first appeared to be only of academic interest, for although (3) provides a consistent test, it is less natural than the tabulated W_N^2. It will be noted that U_N^2 has the form of a variance while W_N^2 has the form of a second moment about the origin, i.e. the modification corresponds to a 'correction for the mean'. Further comparisons are made in § 3.

The above goodness-of-fit tests refer to cases where the sample space is the real line. Turning now to the case where the probability is distributed on the circumference of a circle, a difficulty arises in the application of these tests—there is no natural starting point for the distribution function and different arbitrary starting points will give the test statistics, (1) and (2) above, different values. Kuiper (1960) raises this problem and suggests a variant of (2). He supposes that some arbitrary starting point has been used for a circle of unit circumference. It is clear that

$$V_N = \sup_{0 \leqslant x \leqslant 1} \{F_N(x) - F(x)\} - \inf_{0 \leqslant x \leqslant 1} \{F_N(x) - F(x)\} \tag{4}$$

is independent of the starting point and he shows, using an unpublished result of Darling's, that

$$\text{prob}\{\sqrt{N} V_N \leqslant c\} = 1 - \sum_{j=1}^{\infty} 2(4j^2c^2 - 1) e^{-2j^2c^2} + \frac{8}{3\sqrt{N}} c \sum_{j=1}^{\infty} j^2(4j^2c^2 - 3) e^{-2j^2c^2} + 0\left(\frac{1}{N}\right). \tag{5}$$

† The main result of this paper was derived in 1959 while the author was a Research Associate, Department of Mathematics, Princeton University, and partially supported by ONR funds.

He checks his results against 200 artificial samples of 10. Kuiper also gives a similar solution to the two-sample problem.

The suggestion of this paper is that statistic (3) be used, with the circumference of the circle replacing the real line as the region of integration. For it will be shown in § 2 that its distribution is convenient and does not depend on the starting point. Greenwood (1959) in an unpublished report made the same suggestion but was unable to give the relevant distribution.

When some definite parametric class of alternative distributions is given, neither test (3) nor test (4) should be used since it should be possible to derive a special test that is more powerful than these general purpose tests.

A two-sample test based on an extension of (3) will be given in a further paper and applied there to some real data.

2. Asymptotic distribution of U_N^2

It is first necessary to show that (3), when adapted to the circle of unit circumference is independent of the point chosen for starting to cumulate the distributions. Let $d_x F$ denote the hypothetical probability element at x. Suppose there is another probability distribution on the circle and that $G(x; x_0)$ and $G(x; x_1)$ are its cumulative distribution functions begun respectively at $x = x_0$ and $x = x_1$. All these functions are naturally of period 1.

Taking $0 \leqslant x_1 - x_0 \leqslant 1$, it is clear that

$$\left.\begin{aligned} G(x; x_0) &= G(x; x_1) + G(x_1; x_0), \quad x_1 \leqslant x \leqslant x_0 + 1, \\ &= G(x; x_1) - 1 + G(x_1; x_0), \quad x_0 + 1 \leqslant x \leqslant x_1 + 1. \end{aligned}\right\} \tag{6}$$

Hence
$$G(x; x_0) - \int_0^1 G(x; x_0)\, d_x F = G(x; x_1) - \int_0^1 G(x; x_1)\, d_x F + \int_{x_0}^{x_1} d_x F, \tag{7}$$

where $x_1 \leqslant x \leqslant x_0 + 1$, and for $x_0 + 1 \leqslant x \leqslant x_1 + 1$ the right-hand side of (7) is decreased by one. Thus, if there is yet another distribution on the circle with cumulatives $H(x; x_0)$, $H(x; x_1)$, (7) will again be true with H replacing G. Hence

$$\{G(x; x_0) - H(x; x_0)\} - \int_0^1 \{G(x; x_0) - H(x; x_0)\}\, d_x F$$

is independent of the value of x_0 which may be taken as zero for convenience. Identifying the G-distribution with the sample distribution $F_N(x)$ and the H-distribution with the hypothetical distribution $F(x)$, we may now write, without ambiguity,

$$U_N^2 = N \int_0^1 \left\{ F_N(x) - F(x) - \int_0^1 [F_N(x) - F(x)]\, d_x F \right\}^2 d_x F,$$

where
$$d_x F = dF(x).$$

Since there is no loss of generality in assuming the hypothetical distribution to be uniform, the canonical form of our statistic is

$$U_N^2 = N \int_0^1 \left\{ F_N(u) - u - \int_0^1 [F_N(y) - y]\, dy \right\}^2 du. \tag{8}$$

To find the limiting distribution of (8) we follow the approach of Doob (1949) and define

$$Z_N(u) = \sqrt{N} \left\{ F_N(u) - u - \int_0^1 [F_N(y) - y]\, dy \right\}. \tag{9}$$

For $0 \leqslant u, v \leqslant 1$, it may be shown that

$$\operatorname{cov}\{Z_N(u),\ Z_N(v)\} = \min(u,v) - \tfrac{1}{2}(u+v) + \tfrac{1}{2}(u-v)^2 + \tfrac{1}{12}.$$

Since $\operatorname{var}\{Z_N(u)\} = \frac{1}{12}$, it is convenient to write

$$X_N(u) = \sqrt{12}\, Z_N(u),$$

hence for $-1 \leqslant \tau \leqslant 1$,

$$\operatorname{cov}\{X_N(u),\ X_N(u+\tau)\} = 1 - 6|\tau|(1-|\tau|). \tag{10}$$

Thus

$$12U_N^2 = \int_0^1 X_N(u)^2\, du. \tag{11}$$

As $N \to \infty$, $X_N(u)$ may be taken as a Gaussian process $X(u)$ with zero mean and a covariance function given by the right-hand side of (10) which we will call $\rho(\tau)$.

The asymptotic distribution of $12U_N^2$ may now be derived by the methods of Kac & Siegert (1947*a*). They are merely a generalization of the usual methods for obtaining the distribution of the sum of squares of a finite number of multivariate normal variables with zero means. Thus we need the eigen-solutions of

$$\int_0^1 \rho(s-t)\, h(t)\, dt = \lambda h(s) \quad (0 \leqslant s \leqslant 1). \tag{12}$$

Inspection shows that $h =$ constant, $\lambda = 0$ is a solution. By splitting the range of integration at $t = s$ and differentiating (12) with respect to s, it may be reduced to

$$\left.\begin{aligned} h''(s) + \frac{12}{\lambda} h(s) = 0, \\ \int_0^1 h(s)\, ds = 0. \end{aligned}\right\} \tag{13}$$

The second condition of (13) is required to give orthogonality of all the solutions. Thus we have

$$\left.\begin{aligned} h(s) &= \text{constant} \qquad (\lambda = 0), \\ h(s) &= \cos 2m\pi x \\ h(s) &= \sin 2m\pi x \end{aligned}\right\} \quad \left(\lambda = \frac{3}{m^2\pi^2}\right). \tag{14}$$

Hence $12U_N^2$ is distributed asymptotically as

$$12U^2 = \sum_{m=1}^{\infty} \frac{3}{m^2\pi^2}(a_m^2 + b_m^2), \tag{15}$$

where a_m, $b_m (m = 1, \ldots, \infty)$ are independent sequences of independent standard normal variables. If $w_m (m = 1, \ldots, \infty)$ is a sequence of independent random variables with the same density functions e^{-w}, then

$$U^2 = \sum_{m=1}^{\infty} \frac{1}{2m^2\pi^2} w_m. \tag{16}$$

Thus the moment generating function of U^2 is given by

$$E(e^{\theta U^2}) = \prod_{m=1}^{\infty} \left(1 - \frac{\theta}{2m^2\pi^2}\right)^{-1}, \tag{17}$$

$$= \left[\frac{\sin\sqrt{\tfrac{1}{2}\theta}}{\sqrt{\tfrac{1}{2}\theta}}\right]^{-1}, \tag{18}$$

by the infinite product for the sine function. Thus

$$E(e^{\theta U^2}) = \sum_{m=1}^{\infty} c_m \frac{1}{1 - (\theta/2m^2\pi^2)} \tag{19}$$

where
$$c_m = \lim_{\theta \to 2m^2\pi^2} \frac{1-(\theta/2m^2\pi^2)}{\frac{\sin\sqrt{\frac{1}{2}\theta}}{\sqrt{\frac{1}{2}\theta}}}$$

i.e.
$$c_m = (-1)^{m-1}\,2. \tag{20}$$

Inverting (19) with this value for c_m, the density function of U^2 is found to be

$$\phi(v) = \sum_{m=1}^{\infty} (-1)^{m-1} 2 . 2m^2\pi^2 e^{-2m^2\pi^2 v}. \tag{21}$$

Thus
$$\text{prob}\,(U^2 > v) = \sum_{m=1}^{\infty} (-1)^{m-1} 2 e^{-2m^2\pi^2 v} \tag{22}$$
$$= 2\{e^{-2\pi^2 v} - e^{-8\pi^2 v} + e^{-18\pi^2 v} - \ldots\},$$

which is seen to converge very rapidly. (22) is thus the limiting distribution of U_N^2 as $N \to \infty$. It does not contain, as (5) does, any correction terms for finite n. Both solutions have a form which is characteristic of distributions obtained in this area. The distribution (22) has arisen before in noise theory—see e.g. Kac & Siegert (1947*b*) and Lampard (1956). Even more interesting is the fact that (22) is essentially the asymptotic distribution of K_N^2; to be precise,

$$\lim_{N\to\infty} \text{prob}\,\{K_N < \pi\sqrt{v}\} = \lim_{N\to\infty} \text{prob}\,\{U_N^2 < v\}.$$

It seems very surprising that the asymptotic distributions of $K_N^2\pi^{-2}$ and U_N^2 should be the same. The distribution (22) has been tabulated by Smirnov (1948). The same paper gives an alternative form for (22) which is useful if v is very small.

The reason why statistic (3) is so much easier to handle than statistic (1) is of some interest. It will be recalled that in the distribution theory of the serial correlation coefficients, the circular redefinition of the coefficients introduces circulant matrices and usually, after elimination of the mean, the remaining latent roots are paired. Precisely the same things happen here. The covariance function (10) is, for $0 \leqslant \tau \leqslant 1$, symmetrical about $\tau = \frac{1}{2}$, the odd root is eliminated (i.e. is zero now) and the remainder are paired. There is the further simplification that their pattern is such that the moment generating function is an elementary function. For the case when $F(x)$ involves parameters requiring estimation, there is the double difficulty that the appropriate eigen-values are in general simple (making the inversion difficult) and, in general, extremely hard to find. Again, no way has been found, along these lines, of overcoming these difficulties.

3. Comments on U_N^2

The statistic U_N^2 arose in another connexion, was observed to have a simple distribution and then its relevance to the goodness-of-fit test on the circle was noted. It is possible, however, to introduce U_N^2 in a way that may be more attractive to the reader. Using the notation of the beginning of § 2, suppose we wish to choose rationally some starting point x_0 on the circle in order to make the goodness-of-fit test based on W_N^2. Let us write

$$W_N^2(x_0) = N\int_0^1 \{F_N(x; x_0) - F(x; x_0)\}^2 d_x F. \tag{23}$$

One immediate suggestion would be to choose x_0 to make $W_N^2(x_0)$ a minimum. From (6), with

$$\Delta(x; x_0) = F_N(x; x_0) - F(x; x_0),$$

we derive
$$\Delta(x; x_0) = \Delta(x; x_1) + \Delta(x_1; x_0).$$

Thus
$$\min_{x_0} W_N^2(x_0) = \min_{\Delta} N \int_0^1 \{[F_N(x;0) - F(x;0)] - \Delta\}^2 d_x F, \tag{24}$$

where Δ is a constant. But the right-hand side of (24) is trivially shown to be

$$U_N^2 = N \int_0^1 \left\{F_N(x) - F(x) - \int_0^1 [F_N(y) - F(y)]\, d_y F\right\}^2 d_x F$$

where, as in § 2, $F_N(x) \equiv F_N(x;0)$, $F(x) \equiv F(x;0)$.

Thus
$$U_N^2 = \min_{x_0} W_N^2(x_0).$$

This result is also evident from the computing forms of the statistics. For the *ordered* sample

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(N)}, \text{ write, with } F(x) \equiv F(x;0),$$
$$v_i = F(x_{(i)}).$$

Then
$$W_N^2(0) = \sum_{i=1}^{N} \left(v_i - \frac{2i-1}{2N}\right)^2 + \frac{1}{12N} \tag{25}$$

is the usual form for W_N^2. To find the form for U_N^2 a similar method may be used with the addition of the result that

$$\int_0^1 [F_N(x) - F(x)]\, d_x F = \frac{1}{2} - \frac{1}{N} \sum_{i=1}^{N} F(x_{(i)}),$$
$$= \tfrac{1}{2} - \bar{v}.$$

Hence it may be shown that

$$U_N^2 = \sum_{i=1}^{N} \left(v_i - \frac{2i-1}{2N} - \bar{v} + \tfrac{1}{2}\right)^2 + \frac{1}{12N}. \tag{26}$$

Since $\bar{v} - \frac{1}{2}$ is the arithmetic mean of $v_i - (2i-1)/2N$ $(i = 1, 2, \ldots, N)$, W_N^2 and U_N^2 are again seen as respectively a sum of squares about the origin and of squares of deviations from a mean.

Actually it may be easier to use, instead of (26), the equivalent form

$$U_N^2 = \sum_{i=1}^{N} v_i^2 - 2 \sum_{i=1}^{N} \frac{2i-1}{2N} v_i + \frac{N}{3} + N(\bar{v} - \tfrac{1}{2})^2. \tag{27}$$

Since we are only concerned here with the circular problem, there is no point in a comparison of W_N^2 and U_N^2, but only between U_N^2 and V_N. This latter comparison should be made with respect to their powers but no such information is known. For practical applications it is necessary to use the asymptotic distribution (22), since the distribution of U_N^2 in small samples is not known. The merit of this approximation is untested at present. Assuming that it is satisfactory, the method of computation of U_N^2 from (27), and the derivation of an approximate significance level from (22) are very simple so that no artificial example is called for and the only real data in the author's hands concerns the two-sample problem.

The author is grateful to the Editor for some helpful comments.

REFERENCES

DARLING, D. A. (1957). The Kolmogorov–Smirnov, Cramér–Von Mises tests. *Ann. Math. Statist.* **28**, 823–38.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20**, 393–403.

KAC, M. & SIEGERT, A. J. F. (1947*a*). An explicit representation of a stationary Gaussian process. *Ann. Math. Statist.* **18**, 438–42.

KAC, M. & SIEGERT, A. J. F. (1947*b*). On the theory of noise in radio receivers with square law detectors. *Jour. Appl. Phys.* **18**, 383–97.

KUIPER, N. H. (1960). Tests concerning random points on a circle. *Proc. Koninkl. Nederl. Akad. Van Wettenschappen*, Series A, **63**, 38–47.

LAMPARD, S. G. (1956). The probability distribution for the filtered output of a multiplier. *I.R.E. Trans. Professional Group on Information Theory*, IT-2, 1, 4–11.

SMIRNOV, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.* **19**, 279–81.

Biometrika (1961), **48**, 1 *and* 2, *p.* 115
Printed in Great Britain

The use of orthogonal polynomials of the positive and negative binomial frequency functions in curve fitting by Aitken's method

BY H. T. GONIN
University of South Africa

1. *Introduction.* Frequency data not far from the normal can be fairly represented by a type A series, which often is a normal function plus a corrective term involving the third Hermite polynomial. Frequency data thought to be nearly of Poisson type, but with certain perturbations away from that, can correspondingly be fairly represented by a Poisson function, plus a corrective term involving the second Charlier polynomial for the value of the mean in question, most easily achieved, however, by taking the second backward finite differences of the Poisson fitted class frequencies, multiplying these by a certain easily computed constant, and so correcting the Poisson class frequencies by (usually) small amounts.

In this paper we consider the graduation of a set of data using an expansion similar to the Gram–Charlier type B expansion but based on (i) the positive binomial, (ii) the negative binomial. The form thus assumed is $y_x = \{a_0 + a_1 G_1(x) + a_2 G_2(x) + \ldots\}\phi(x)$, where the G's are an orthogonal set with respect to $\phi(x)$ (which is binomial) and where the problem is to determine 'best' values of $a_0, a_1, \ldots$, etc. The index n and the fraction p of the binomial are assumed known, or are estimated from the first two moments of the data. We are mainly concerned with a fit to a set of data of nearly binomial form, and the problem of efficiently estimating the parameters for a given Gram–Charlier type B model is more complicated and of a somewhat different nature.

The method used is that of weighted least squares, the weights being $\phi(x)^{-1}$. In the case of many data, such as, e.g. the number of deaths θ_x in a mortality table, the weights are proportional to the inverse of the square of the standard deviation and approximate to y_x^{-1} (in our case we use $\phi(x)$ as a first approximation to y_x). The scheme followed is that introduced by Aitken (1932–33) in the unweighted rectangular case where the orthogonal polynomials are those of Tchebycheff.

In the latter case Fisher's scheme is perhaps just as simple, but the tabulation of the polynomials in the binomial case introduces a serious complication. The values of the orthogonal polynomials would have to be tabulated, in the case of the positive binomial, e.g. for two parameters, the first being the value of p (and this, for fine grading, would need hundreds of values), the second the number of data. By using Aitken's method the corrective terms emerge as simple by-products of the factorial moments and so the work is quite economical.

2. *Aitken's method of fitting in the weighted case.* Let us fit a series of observed frequencies O_x to the curve

$$y_x = \{a_0 G_0(x) + a_1 G_1(x) + \ldots + a_r G_r(x)\}\phi(x) \qquad (1)$$

by the method of least squares, with weighting function $\phi(x)^{-1}$.

By Legendre's principle

$$S^2 = \sum_x \phi(x)^{-1}[O_x - \{a_0 G_0 + a_1 G_1 + \ldots + a_r G_r\}\phi(x)]^2$$

must be a minimum.

Hence $\partial S^2/\partial a_r = 0$. Thus

$$a_r \sum_x \phi(x)\, G_r^2(x) = \sum_x O_x G_r(x). \tag{2}$$

Let
$$G_r(x) = b_0^r + b_1^r x_{(1)} + b_2^r x_{(2)} + \ldots + b_r^r x_{(r)}, \tag{3}$$

where
$$x_{(r)} = \frac{x^{(r)}}{r!} \quad \text{and} \quad x^{(r)} = x(x-1)\ldots(x-r+1).$$

We multiply (3) by O_x and sum over all the values of x, obtaining

$$\sum_x O_x G_r(x) = b_0^r \sum_x O_x + b_1^r \sum_x O_x x_{(1)} + \ldots + b_r^r \sum_x O_x x_{(r)}$$

$$= b_0^r m_{(0)} + b_1^r m_{(1)} + \ldots + b_r^r m_{(r)}, \tag{4}$$

where $m_{(r)} = \Sigma O_x x_{(r)}$, the rth reduced factorial moment of the data.

Now (3) is really a Gregory–Newton interpolation formula. Hence the coefficients of $1, x_{(1)}, \ldots, x_{(r)}$ of the polynomial $G_r(x)$, namely $b_0^r, b_1^r, \ldots, b_r^r$, are the terminal values $G_r(0)$, $\Delta G_r(0), \ldots, \Delta^r G_r(0)$. (The indices of the coefficients indicate to what degree of polynomial they belong, thus b_r^s denotes the coefficients of $x_{(r)}$ in $G_s(x)$.)

This simple fact, noticed and turned to use by Aitken, is of great computational value in his scheme. From (2) and (4) we have

$$a_r \sum_x \phi(x)\, G_r^2(x) = b_0^r m_{(0)} + b_1^r m_{(1)} + \ldots + b_r^r m_{(r)}. \tag{5}$$

Hence the constants a_r can be computed by combining the various coefficients of $1, x_{(1)}, \ldots, x_{(r)}$ of our polynomials with the reduced factorial moments of the observed frequencies and dividing by $\sum_x \phi(x)\, G_r^2(x)$.

In the Aitken table the $(j+1)$th column will be $b_0^j, b_1^j, \ldots, b_j^j$ and the $(i+1)$th row will be $b_i^i, b_i^{i+1}, \ldots, b_i^r$, if constants $a_0, a_1, \ldots, a_r$ are required. The $(j+1)$th column therefore merely consists of the coefficients of $G_j(x)$ written in the form of reduced factorials, the column starting with the constant.

Again we have from equation (1) that

$$\frac{y_0}{\phi(0)} = a_0 G_0(0) + a_1 G_1(0) + a_2 G_2(0) + \ldots + a_r G_r(0),$$

$$\Delta \frac{y_0}{\phi(0)} = \qquad a_1 \Delta G_1(0) + a_2 \Delta G_2(0) + \ldots + a_r \Delta G_r(0),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\Delta^r \frac{y_0}{\phi(0)} = \qquad\qquad\qquad\qquad\qquad\qquad a_r \Delta^r G_r(0),$$

or

$$\frac{y_0}{\phi(0)} = a_0 b_0^0 + a_1 b_0^1 + a_2 b_0^2 + \ldots + a_r b_0^r,$$

$$\Delta \frac{y_0}{\phi(0)} = \qquad a_1 b_1^1 + a_2 b_1^2 + \ldots + a_r b_1^r,$$

$$\Delta^r \frac{y_0}{\phi(0)} = \qquad\qquad\qquad\qquad a_r b_r^r.$$

The whole graduation therefore falls into four easy steps:

(1) Compute the reduced factorial moments by repeated summation.

(2) Combine these according to multipliers in the columns of coefficients of $1, x_{(1)}, \ldots, x_{(r)}$ of the orthogonal polynomials $G_r(x)$ and divide by $\Sigma\phi(x)\,G_r^2(x)$, the values set out at the feet of the columns.

(3) Combine the a_r values obtained by step (2) according to multipliers in the rows of the same table, thus obtaining $y_0/\phi(0)$, $\Delta[y_0/\phi(0)]$, ..., $\Delta^r[y_0/\phi(0)]$.

(4) Complete the difference table by summation to obtain the other graduated values of $y_x/\phi(x)$.

3. *The orthogonal polynomials of the positive and negative binomials.* For the positive binomial $\binom{n}{x}p^xq^{n-x}$, the orthogonal polynomials are given by

$$G_r(x;\,n,p) = (1+p\Delta)^{-n+r-1}x^{(r)}, \quad \text{symbolically}$$
$$= x^{(r)} - r(n-r+1)\,p\,x^{(r-1)} + \binom{r}{2}(n-r+2)^{(2)}p^2x^{(r-2)} + \ldots + (-1)^r n^{(r)}p^r, \qquad (6)$$

where $x^{(r)} = x(x-1)\ldots(x-r+1)$ (Aitken & Gonin, 1935).

The explicit form for the polynomials of the negative binomial

$$\frac{(\alpha)_x}{x!}\left(\frac{\lambda}{\alpha}\right)^x\left(1+\frac{\lambda}{\alpha}\right)^{-\alpha-x}, \quad \text{where} \quad (\alpha)_x = \alpha(\alpha+1)\ldots(\alpha+x-1) = (\alpha+x-1)^{(x)}$$

was first given by Shenton (1950) and the symbolic form by the same author in 1958. The polynomials were also fully studied by Wiid (1958). They are, in Shenton's notation, given by

$$G_r(x;\,\lambda,\alpha) = \left(1-\frac{\lambda\Delta}{\alpha}\right)^{\alpha+r-1}x^{(r)} \quad (\lambda \text{ and } \alpha > 0)$$
$$= x^{(r)} - r(\alpha+r-1)\frac{\lambda}{\alpha}x^{(r-1)} + \binom{r}{2}(\alpha+r-1)^{(2)}\frac{\lambda^2}{\alpha^2}x^{(r-2)} + \ldots + (-1)^r(\alpha+r-1)^{(r)}\frac{\lambda^r}{\alpha^r}. \qquad (7)$$

All the relationships between the frequency functions, polynomials, factorial moments, generating functions, etc., for the negative binomial can be derived from the positive binomial by putting $p = -\lambda/\alpha$ and $n = -\alpha$, e.g.

		Positive binomial	Negative binomial
(*a*) Factorial moments	$\sum_x \phi(x)\,x^{(r)}$	$n^{(r)}p^r$	$(\alpha)_r\frac{\lambda^r}{\alpha^r}$
(*b*)	$\sum_x \phi(x)\,G_r^2(x)$	$r!\,n^{(r)}p^rq^r$	$r!\,(\alpha)_r\frac{\lambda^r}{\alpha^r}\left(1+\frac{\lambda}{\alpha}\right)^r$
(*c*) Generating function	$\sum_r G_r(x)\frac{t^r}{r!}$	$(1+qt)^x(1-pt)^{n-x}$	$\left(1+1+\frac{\lambda}{\alpha}t\right)^x\left(1+\frac{\lambda}{\alpha}t\right)^{-\alpha-x}$

4. *Fitting by the polynomials of the positive binomial.* (*a*) A fitting by the symmetrical polynomials where $p = q = \frac{1}{2}$ was done by Greenleaf (1932) and Gonin (1945), the latter using Aitken's method which proved to be much quicker than Greenleaf's, especially when the polynomials were expressed in terms of central and mean central factorials. Now if we change the polynomials given by (6) in such a way that the constant term is unity, and denote these polynomials by $G'_r(x;\,n,p)$, we can easily prove that

$$\binom{n}{x}G'_r(x;\,n,\tfrac{1}{2}) = H_r^n(x), \quad \text{where} \quad H_r^n(x) = \Delta^r\binom{n-r}{x-r},$$

the elegant orthogonal functions introduced by Milne and Stone (see Milne, 1949) and used by them for graduating data. The function $H_r^n(x)$ has this peculiar orthogonal property (in Milne's notation)

$$\sum_{k=0}^{n} H_k^n(s)\, H_l^n(k) = 0 \quad \text{if} \quad s \neq t,$$

$$= 2^n \quad \text{if} \quad s = t$$

and the recurrence relationship

$$H_{l+1}(s+1) = H_l(s+1) - H_{l+1}(s) - H_l(s).$$

A table of $H_r^n(x)$ is easily computed, the first row in the table being binomial coefficients and the first column being all unity. By using the recurrence formula the table is quickly completed and can be used following Fisher's method, since if we graduate by

$$y_x = \{a_0 G_0'(x) + a_1 G_1'(x) + \ldots + a_r G_r'(x)\} \binom{n}{x} (\tfrac{1}{2})^x (\tfrac{1}{2})^{n-x}$$

this is equivalent to graduating by

$$y_x = \frac{1}{2^n}\{a_0 H_0^n(x) + a_1 H_1^n(x) + \ldots + a_r H_r^n(x)\}.$$

(*b*) If $p \neq q$, it is best to graduate by Aitken's method. Let us graduate Weldon's data in which he records the fives and sixes in 26,306 throws of 12 dice. These data have been graduated by Fry (1928) by various curves.

We first compute the reduced factorial moments in the usual way by summing from the bottoms of the respective columns, stopping one item from the top each time, obtaining $m_{(r)} = \sum_x O_x x_{(r)}$. We obtain

$$m_{(0)} = 26{,}306, \quad m_{(1)} = 106{,}602, \quad m_{(2)} = 198{,}184, \quad m_{(3)} = 223{,}524.$$

The values are entered at the side of Aitken's table. The first three polynomials for $n = 12$ and $p = \frac{1}{3}$ are $x - 4$, $2x_{(2)} - \frac{22}{3}x + \frac{44}{3}$ and $6x_{(3)} - 20x_{(2)} + \frac{110}{3}x - \frac{440}{9}$.

The values of $\sum_x \phi(x)\, G_r^2(x)$ are entered at the feet of the columns

		r	0	1	2	3
		a_r	26,306	516·75	33·8523	1·5290
$m_{(0)}$ 26,306	$G_r(0)$	.	1	−4	$\frac{44}{3}$	$-\frac{440}{9}$
$m_{(1)}$ 106,602	$\Delta G_r(0)$	.	.	1	$-\frac{22}{3}$	$\frac{110}{3}$
$m_{(2)}$ 198,184	$\Delta^2 G_r(0)$	.	.	.	2	−20
$m_{(3)}$ 223,524	$\Delta^3 G_r(0)$	.	.	.	.	6
$\Sigma\phi(x)\, G_r^2(x)$			1	$\frac{8}{3}$	$\frac{352}{27}$	$\frac{7040}{81}$

$$\begin{cases} a_0 = 26{,}306, \\ a_1 = (106{,}602 - 4 \times 26{,}306) \div \frac{8}{3} = 516{\cdot}75, \\ a_2 = 198{,}184(2) + 106{,}602(-\frac{22}{3}) + 26{,}306(\frac{44}{3})/\frac{352}{27}, \\ \quad = 33{\cdot}8523, \\ a_3 = 223{,}524(6) + 198{,}184(-20) + 106{,}602(\frac{110}{3}) + 26{,}306(-\frac{440}{9})/\frac{7040}{81}, \\ \quad = 1{\cdot}5290. \end{cases}$$

Therefore

$$\begin{cases} \dfrac{y_0}{\phi(0)} = a_0 + a_1 G_1(0) + a_2 G_2(0) + a_3 G_3(0) \\ \qquad = 26{,}306 + 516{\cdot}75(-4) + 33{\cdot}8523(\tfrac{44}{3}) + 1{\cdot}5290(-\tfrac{440}{9}) \\ \qquad = 24{,}660{\cdot}75, \\ \Delta \dfrac{y_0}{\phi(0)} = 516{\cdot}75 - 33{\cdot}8523(\tfrac{22}{3}) + 1{\cdot}5290(\tfrac{110}{3}) \\ \qquad = 324{\cdot}56, \\ \Delta^2 \dfrac{y_0}{\phi(0)} = 37{\cdot}12, \\ \Delta^3 \dfrac{y_0}{\phi(0)} = 9{\cdot}17. \end{cases}$$

A good check on the polynomials is to enter the theoretical reduced factorial moments $n^{(r)}p^r/r!$ at the side of the table instead of the $m'_{(r)}$, then all the a_r values must be zero except a_0.

By completing the difference table we obtain $y_x/\phi(x)$ and multiplying by $\phi(x)$ we obtain y_x.

x	$\frac{y_x}{\phi(x)}$	y_x (graduated)	Empirical graduation by Fry	Observed data O_x	Graduation by $\phi(x)$ alone
0	24,661	190	190	185	203
1	24,985	1,155	1,155	1,149	1,217
2	25,347	3,224	3,223	3,265	3,345
3	25,755	5,461	5,459	5,475	5,576
4	26,218	6,252	6,252	6,114	6,273
5	26,746	5,100	5,102	5,194	5,018
6	27,348	3,041	3,043	3,067	2,927
7	28,033	1,336	1,337	1,331	1,255
8	28,810	430	429	403	392
9	29,688	99	98	105	87
10	30,677	16	15	14	13
11	31,786	1·6	1·4	4	1
12	33,623	0·1	0·1	0	0

There is a very close agreement with the empirical graduation of Fry, in which he uses the formula

$$\phi(x) = \binom{12{\cdot}135}{x} (0{\cdot}339325)^x (0{\cdot}660675)^{12{\cdot}135-x}.$$

There is an objection, mentioned by Fry, to the empirical fitting since we know that n is 12 and not 12·135. His fitting is, however, better than ours as he has one more degree of freedom in the χ^2 test. Our fitting is, nevertheless, very satisfactory and a great improvement on the graduation by $\phi(x)$ alone.

5. *Graduation by polynomials of the negative binomial.* We graduate by

$$y_x = [a_0 + a_1 G_1(x) + \ldots + a_r G_r(x)]\,\phi(x),$$

where $\phi(x)$ is the negative binomial frequency function already defined.

The a_r values are obtained from the relationship

$$\sum_x G_r(x)\, O_x = a_r \sum_x \phi(x)\, G_r^2(x),$$

where $$\Sigma\phi(x)\,G_r^2(x) = r!(\alpha)_r \frac{\lambda^r(\lambda+\alpha)^r}{\alpha^{2r}}$$

and $$\lambda = \frac{NM_{(1)}}{N} = \bar{x}, \quad \alpha = \frac{\bar{x}^2}{s^2-\bar{x}},$$

using moment estimators, where $\bar{x}$ is the mean of the distribution and s^2 is the second moment with zero origin.

We choose an example based on the data given by Hald (1952, p. 730) and graduated by him by means of a negative binomial. The reduced factorial moments are first calculated.

Distribution of 647 *women according to number of accidents.*

(The values in bold figures are the reduced factorial moments)

0	447	**647**			
1	132	200	**301**		
2	42	68	101	**143**	
3	21	26	33	42	**53**
4	3	5	7	9	11
5	2	2	2	2	2
6	0	0	0	0	0

$\lambda = \bar{x} = \frac{301}{647} = 0{\cdot}4652$, $s^2 = 0{\cdot}6919$, agreeing with Hald.

Therefore $$\alpha = 0{\cdot}9546 \quad \text{and} \quad \lambda/\alpha = 0{\cdot}4873.$$

Since the parameters are estimated in this way, it follows that a_1 and a_2 will both be zero and we need only compute $G_3(x)$ and $\sum_x \phi(x)\,G_3^2(x)$.

$$\begin{aligned} G_3(x) &= 6x_{(3)} - 6(2{\cdot}9546)\,(0{\cdot}4873)\,x_{(2)} + 3(2{\cdot}9546)^{(2)}\,(0{\cdot}4873)^2\,x_{(1)} \\ &\quad - (2{\cdot}9546)^{(3)}\,(0{\cdot}4873)^3 \\ &= 6x_{(3)} - 8{\cdot}6387x_{(2)} + 4{\cdot}1140x_{(1)} - 0{\cdot}6378. \end{aligned}$$

$$\sum_x \phi(x)\,G_3^2(x) = 3!\,(2{\cdot}9546)^{(3)}\,(0{\cdot}4873)^3\,(1{\cdot}4873)^3 = 12{\cdot}5902.$$

Hence our table becomes

	r	0	1	2	3
	a_r	647	.	.	−7·2816
647		.	1	.	.	−0·6378
$m_{(r)}$ 301		.	.	.	.	4·1140
143		.	.	.	.	−8·6387
53		.	.	.	.	6
$\Sigma\phi(x)\,G_r^2(x)$		.	1	.	.	12·5902

The differences are

$$\frac{y_0}{\phi(0)} = 652, \quad \Delta\frac{y_0}{\phi(0)} = -30, \quad \Delta^2\frac{y_0}{\phi(0)} = 63, \quad \Delta^3\frac{y_0}{\phi(0)} = -44.$$

By building up a difference table we obtain the values $y_x/\phi(x)$. Or we can use

$$\frac{y_x}{\phi(x)} = 652 - x30 + \frac{x(x-1)}{2!}63 - \frac{x(x-1)\,(x-2)}{3!}44.$$

The graduation results are

x	O_x(observed)	$\frac{y_x}{\phi(x)}$	y_x (graduated)	χ^2	Hald (negative binomial)	χ^2
0	447	652	446·2	0·0	442·9	0·4
1	132	622	133·1	0·0	138·5	0·3
2	42	655	44·9	0·2	44·4	0·1
3	21	707	15·6	1·9	14·3	3·1
4	3	734	5·2	0·9	4·6	0·6
5	2	692	1·6	0·1	1·5	0·2
$\geqslant 6$	0	537	0·6	0·6	0·7	0·7
Totals	647	.	647·2	2·7	646·9	5·4
			D.F. = 3		D.F. = 4	
			$P = 0{\cdot}45$		$P = 0{\cdot}25$	

The graduation is obviously excellent and an improvement on Hald's, even allowing for the fact that there is one degree of freedom less. The parameters were also calculated by the method of maximum likelihood, using the method of Haldane (1941), and the corresponding polynomials constructed. The results showed little improvement on that of our method and we lost the advantage of having a_1 and a_2 zero, so that all the polynomials had to be constructed.

We examined about five other distributions which were graduated by the negative binomial. In every case the use of the third polynomial brings about a good improvement in the graduation. In the positive binomial case the integer value of n is usually known and this should be used and not that obtained from the method of moment estimators.

Shenton has another method of estimating the parameters λ and α by using orthogonal polynomials up to the third degree for $\partial\phi/\partial\lambda$ and $\partial\phi/\partial\alpha$ This involves much arithmetic and our method of using the third polynomials with moment estimators has more or less the same effect.

There is another objection which is often brought forward against a graduation by a series of polynomials. It is said that meaning of the parameters such as they occur, e.g. in ecology, are disturbed. The question arises whether such a graduating series is entirely empirical or is there a justification for its employment. In the case of the positive and negative binomial there is such a justification. Consider the frequency function $\binom{n}{x} p^x q^{n-x}$. Let p' and q' be the estimated parameters and p and q the true parameters. Then if p' and q' deviate slightly from the true parameters, we have $p = p' + \epsilon$ and $q = q' - \epsilon$, where ϵ is a small quantity such as 0·005, which may, for example, be caused by wear and tear of the dice. Hence

$$\begin{aligned}\binom{n}{x} p^x q^{n-x} &= \binom{n}{x} (p'+\epsilon)^x (q'-\epsilon)^{(n-x)} \\ &= \binom{n}{x} p'^x q'^{n-x} \left(1+\frac{\epsilon}{p'}\right)^x \left(1-\frac{\epsilon}{q'}\right)^{(n-x)} \\ &= \binom{n}{x} p'^x q'^{n-x} \left(1+\frac{\epsilon q'}{p'q'}\right)^x \left(1-\frac{\epsilon p'}{p'q'}\right)^{(n-x)}.\end{aligned}$$

But

$$(1+q't)^x (1-p't)^{(n-x)} = \sum_{r=0}^{n} G_r(x, p') \frac{t^r}{r!},$$

(see Szegö, 1939), where $G_r(x, p')$ is the rth orthogonal polynomial of the weighting function $\binom{n}{x} p'^x q'^{(n-x)}$. Hence

$$\binom{n}{x} p^x q^{n-x} = \binom{n}{x} p'^x q'^{n-x} \left\{ 1 + \frac{\epsilon}{p'q'} G_1(x, p') + \frac{\epsilon^2}{2! p'^2 q'^2} G_2(x, p') + \ldots + \frac{\epsilon^n}{n! p'^n q'^n} G_n(x, p') \right\}.$$

If ϵ is small and p and q are finite values of the order 1, then the coefficients $\epsilon^r / r! p'^r q'^r$ are small and we may find that the series converges quickly and that the use of polynomials up to the 3rd degree already give a very good approximation.

6. *Construction of three polynomials in discrete cases where explicit forms for the polynomials are not available.* The author has found it convenient to express the polynomials in factorials of x instead of powers of x. If $\phi(x)$ is the frequency function, the rth polynomial is given by

$$G_r(x) = \begin{vmatrix} 1 & x^{(1)} & \ldots & x^{(r)} \\ \Sigma\phi(x)\, x^{(0)} x^{(0)} & \Sigma\phi(x)\, x^{(0)} x^{(1)} & \ldots & \Sigma\phi(x)\, x^{(0)} x^{(r)} \\ \Sigma\phi(x)\, x^{(1)} x^{(0)} & \Sigma\phi(x)\, x^{(1)} x^{(1)} & \ldots & \Sigma\phi(x)\, x^{(1)} x^{(r)} \\ \Sigma\phi(x)\, x^{(r-1)} x^{(0)} & \Sigma\phi(x)\, x^{(r-1)} x^{(1)} & \ldots & \Sigma\phi(x)\, x^{(r-1)} x^{(r)} \end{vmatrix} \div P^{(r-1)},$$

where $$P^{(r)} = |\mu_{(st)}| \quad (s = 0, 1, \ldots, r;\ t = 0, 1, \ldots, r)$$

and
$$\begin{aligned} \mu_{(rs)} &= \sum_x \phi(x)\, x^{(r)} x^{(s)} \\ &= \sum_x \phi(x)\, x^{(r)} (\overline{x - r} + r)^{(s)} \\ &= \sum_x \phi(x)\, x^{(r+s)} + r.s \sum_x \phi(x)\, x^{(r+s-1)} + \binom{r}{2} s^{(2)} \sum_x \phi(x)\, x^{(r+s-2)} \ldots \end{aligned}$$

(using Vandermonde's expansion)

$$= \mu_{(r+s)} + r.s\mu_{(r+s-1)} + \binom{r}{2} s^{(2)} \mu_{(r+s-2)} + \ldots.$$

This expansion of moments of the type $\sum_x \phi(x)\, x^{(r)} x^{(s)}$ into a series involving factorial moments has been given by Guest (1953).

I should like to thank Prof. A. J. B. Wiid, lately of the Council of Science and Industrial Research, for his valuable assistance in drawing up this paper, and Mrs H. E. Rudolf, also of the C.S.I.R., for her technical assistance. Further, I wish particularly to thank the referees who, through their helpful remarks, have suggested an appropriate introduction to the paper and also other improvements.

REFERENCES

AITKEN, A. C. (1932–33). On the graduation of data by the orthogonal polynomials of least squares. *Proc. Roy. Soc. Edinb.* **53**, 54–78.

AITKEN, A. C. & GONIN, H. T. (1935). On fourfold sampling with and without replacement. *Proc. Roy. Soc. Edinb.* **55**, 114–25.

FRY, T. C. (1928). *Probability and its Engineering Uses.* New York: D. van Nostrand Co. Inc.

GUEST, P. (1953). The Doolittle method and the fitting of polynomials to weighted data. *Biometrika*, **40**, 229–31.

GONIN, H. T. (1945). Curve fitting by means of the orthogonal polynomials in binomial statistical distributions. *Trans. Roy. Soc. S. Afr.* **30**, 207–15.

GREENLEAF, H. E. H. (1932). Curve approximation by means of functions analogous to the Hermite polynomials. *Ann. Math. Statist.* **3**, 204–55.

HALD, A. (1952). *Statistical Theory with Engineering Applications*. New York: John Wiley and Sons, Inc.

HALDANE, J. B. S. (1941). The fittings of binomial distributions. *Ann. Eugen., Lond.*, **11**, 179–81.

MILNE, W. E. (1949). *Numerical Analysis*. Princeton University Press.

SHENTON, L. R. (1950). Maximum likelihood and the efficiency of the method of moments. *Biometrika*, **37**, 111–16.

SHENTON, L. R. (1958). Moment estimators and maximum likelihood. *Biometrika*, **45**, 411–20.

SZEGÖ, G. (1939). *Orthogonal Polynomials*. New York: Mathematical Society Colloquium Publications, **23**.

WIID, A. J. B. (1958). Afleiding van die ortogonale polinome van die negatiewe binomiale en negatiewe faktoriaalbinomiale verdelings. *Tydskr. Wet. Kuns. Bloemfontein*. Nuwe Reeks Deel **18**.

Biometrika (1961), **48**, 1 *and* 2, *p.* 125
Printed in Great Britain

The estimation of regression and error-scale parameters, when the joint distribution of the errors is of any continuous form and known apart from a scale parameter

BY A. M. W. VERHAGEN

Division of Mathematical Statistics, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia

INTRODUCTION AND SUMMARY

Pitman (1938) discussed the estimation of location and scale parameters of a continuous population of any given form, when n independent observations are available.

In this paper, Pitman's estimation theory and his notions, 'proper region', 'fiducial function' and 'the estimator property' are extended to include the situation where all regression parameters and an error-scale parameter in multiple linear regression models are to be estimated. A further generalization is that the errors need not have independent or identical distributions. It is shown that the joint-fiducial functions may be used to give joint and individual confidence regions for the parameters. 'Closest' and 'least-mean-square-error' estimators for the parameters are expressed in terms of the fiducial function and the general theory developed is illustrated with the case of normal independently distributed errors.

PROBABILITY RELATIONS BETWEEN OBSERVATIONS AND PARAMETERS

Fiducial functions

Consider the regression model $$\mathbf{y} = \mathbf{X\beta} + \sigma\mathbf{u},$$

where X is the $n \times r$ regression matrix of rank r, $\boldsymbol{\beta}$ is the vector $(\beta_1, \ldots, \beta_r)$ of unknown regression coefficients, σ is the unknown positive error-scale parameter. The co-ordinates $u_1, \ldots, u_n$ of the vector $\mathbf{u}$ are assumed to have a completely known joint distribution $F(\mathbf{u})\,d\mathbf{u}$.

For any given vector-set P such that

$$\iiint_P F(\mathbf{u})\,d\mathbf{u} = p$$

the relation $$\frac{\mathbf{y} - \mathbf{X\beta}}{\sigma} \subset P \tag{1}$$

between the random variables $y_1, \ldots, y_n$ and the unknown parameters $\beta_1, \beta_2, \ldots, \beta_r$ and σ is satisfied with probability p irrespective of the true value of the parameters.

The relation (1) will be studied with the aid of subsets $A(\mathbf{y})$ defined as

$$A(\mathbf{y}) \equiv \left\{\frac{\mathbf{y} - \mathbf{Xb}}{s} : 0 < s < \infty, -\infty < b_\nu < \infty, \nu = 1, \ldots, r\right\},^{\dagger} \tag{2}$$

where s takes all values in the range $(0, \infty)$ and the b_ν $(\nu = 1, \ldots, r)$ take all values in the range $(-\infty, \infty)$. Each vector $\mathbf{y}$ has such a set $A(\mathbf{y})$ associated with it, but vectors $\mathbf{y}$ in $A(\mathbf{y})$

† The brace notation $\{:\}$ indicates the set of vectors before the double point, satisfying the conditions behind it.

all lead to the same set $A(\mathbf{y})$ (A is mnemonic for association). The sets $A(\mathbf{y})$ define a subdivision of the space into mutually exclusive sets, and this subdivision defines a subdivision of P into mutually exclusive sets of the form $P \cap A(\mathbf{y})$.

The (unconditional) probability that (1) is satisfied may therefore be expressed in terms of conditional probabilities given the spaces $A(\mathbf{y})$. To bring out the relationship between (1) and the conditional probabilities given $A(\mathbf{y})$ explicitly, a change of variables is needed such that some of the new variables describe the vectors in $A(\mathbf{y})$ and the others are constant within $A(\mathbf{y})$. For this purpose the definition of $A(\mathbf{y})$ suggests writing $\mathbf{y}$ as a sum of two vectors, one in the space spanned by the columns of the matrix $\mathbf{X}$ and one in the space spanned by the columns of a matrix $\widetilde{\mathbf{X}}$. (The square matrix has rank n, $\mathbf{X}$ has rank r and $\widetilde{\mathbf{X}}$ has rank $n-r$.)

This gives

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = [\mathbf{X} \vdots \mathbf{0}] \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ \vdots \\ z_n \end{bmatrix} + [\mathbf{0} \vdots \widetilde{\mathbf{X}}] \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ \vdots \\ z_n \end{bmatrix} = [\mathbf{X} \vdots \widetilde{\mathbf{X}}]\,\mathbf{z}, \tag{3}$$

where $z_1, \ldots, z_r$ and $z_{r+1}, \ldots, z_n$ are co-ordinates of $\mathbf{y}$ with respect to the columns of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$. Hence

$$\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma} = [\mathbf{X} \vdots \widetilde{\mathbf{X}}] \begin{bmatrix} \frac{z_1-\beta_1}{\sigma} \\ \vdots \\ \frac{z_r-\beta_r}{\sigma} \\ \frac{z_{r+1}}{\sigma} \\ \vdots \\ \frac{z_n}{\sigma} \end{bmatrix} = [\mathbf{X} \vdots \widetilde{\mathbf{X}}] \left(\frac{\mathbf{z}-\boldsymbol{\beta}}{\sigma} \right).$$

The change of variables from $\mathbf{y}$ to $\mathbf{z}$ has Jacobian $|\mathbf{X} \vdots \widetilde{\mathbf{X}}| \neq 0$. Let $\mathbf{y}^0$ stand for a definite vector $\mathbf{y}$ and $\mathbf{z}^0$ its transform according to (3). Then the change of variables from $z_1, \ldots, z_n$ to $s, b_1, \ldots, b_r, w_1, \ldots, w_{n-r-1}$ defined by

$$\left.\begin{aligned}
\frac{z_1-\beta_1}{\sigma} &= \frac{z_1^0-b_1}{s}, \\
&\vdots \\
\frac{z_r-\beta_r}{\sigma} &= \frac{z_r^0-b_r}{s}, \\
\frac{z_{r+1}}{\sigma} &= \frac{z_{r+1}^0}{s} = \frac{R}{s}\cos w_1, \\
&\cdots\cdots\cdots \\
\frac{z_{n-1}}{\sigma} &= \frac{z_{n-1}^0}{s} = \frac{R}{s}\sin w_1 \sin w_2 \ldots \cos w_{n-r-1}, \\
\frac{z_n}{\sigma} &= \frac{z_n^0}{s} = \frac{R}{s}\sin w_1 \sin w_2 \ldots \sin w_{n-r-1}
\end{aligned}\right\} \tag{4}$$

is such that the new variables $b_1, \ldots, b_r$ and s describe variations within $A(\mathbf{y}^0)$ and $w_1, \ldots, w_{n-r-1}$ are constant inside $A(\mathbf{y}^0)$. The required transformation has been achieved in two steps and the Jacobian is $\psi(\mathbf{w})\,(1/s)^{n+1}$ ($\mathbf{w}$ stands for $w_1, \ldots, w_{n-2}$). For $\mathbf{w}$ constant, i.e. $\mathbf{y} \subset A(\mathbf{y}^0)$, the change of variable (4) may be written

$$\frac{\mathbf{z}-\boldsymbol{\beta}}{\sigma} = \frac{\mathbf{z}^0-\mathbf{b}}{s} \tag{5}$$

which with the aid of (3) becomes

$$\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma} = \frac{\mathbf{y}^0-\mathbf{X}\mathbf{b}}{s}. \tag{6}$$

Hence the integral

$$\iiint_{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/\sigma \subset P} F\left(\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) d\left(\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \tag{7}$$

transforms to

$$\iiint_{w_1, \ldots, w_{n-r-1}} \frac{\psi(\mathbf{w})}{K(\mathbf{w})} \left[\iiint_{(\mathbf{y}^0-\mathbf{X}\mathbf{b})/s \subset P \cap A(\mathbf{y}^0)} \frac{K(\mathbf{w})}{s^{n+1}} F\left(\frac{\mathbf{y}^0-\mathbf{X}\mathbf{b}}{s}\right) db_1, \ldots, db_r ds\right] dw_1 \ldots dw_{n-r-1}, \tag{8}$$

where $K(\mathbf{w})$ stands for

$$\left[\iiint_{(\mathbf{y}^0-\mathbf{X}\mathbf{b})/s \subset A(\mathbf{y}^0)} \frac{1}{s^{n+1}} F\left(\frac{\mathbf{y}^0-\mathbf{X}\mathbf{b}}{s}\right) db_1, \ldots, db_r, ds\right]^{-1}. \tag{9}$$

Formula (8) gives the probability that (1) is satisfied, and the inner integral in (8) gives the conditional probability that (1) is satisfied given $\mathbf{y} \subset A(\mathbf{y}^0)$.

Probability relations between parameters and observations for the case of n independent observations from a population of the form $f((y-\beta_1)/\sigma)$ in which β_1 is an unknown location parameter and σ an unknown scale parameter were studied by Pitman (1938). It becomes a special case of the theory considered here by taking $r = 1$ and

$$F = f\left(\frac{y_1-\beta_1}{\sigma}\right) f\left(\frac{y_2-\beta_1}{\sigma}\right) \ldots f\left(\frac{y_n-\beta_1}{\sigma}\right)$$

(i.e. errors are independent and have identical distributions).

The function

$$\frac{K(\mathbf{w})}{s^{n+1}} F\left(\frac{\mathbf{y}^0-\mathbf{X}\mathbf{b}}{s}\right) \tag{10}$$

in the inner integral of (8) then reduces to

$$\frac{K(\mathbf{w})}{s^{n+1}} f\left(\frac{y_1^0-b_1}{s}\right) f\left(\frac{y_2^0-b_1}{s}\right) \ldots f\left(\frac{y_n^0-b_1}{s}\right). \tag{11}$$

Pitman (1938) calls (11) the fiducial function for β_1 and σ. By analogy (10) may be called the fiducial function for $\beta_1, \ldots, \beta_r$ and σ.

Proper regions

The integral $I(\mathbf{y}^0)$ of the fiducial function corresponding to an observed vector $\mathbf{y}^0$ over a region of the type

$$\frac{\mathbf{y}^0-\mathbf{X}\mathbf{b}}{s} \subset P \cap A(\mathbf{y}^0) \tag{12}$$

gives the conditional probability that

$$\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma} \subset P \cap A(\mathbf{y}^0) \tag{12a}$$

given $\mathbf{y} \subset A(\mathbf{y}^0)$.

Thus with integration regions of the form (12) it is possible to evaluate the conditional probability of relations of the type (12a) being satisfied. Regions of the form (12) are called 'proper regions' a term borrowed from Pitman (1938). If P is such that the conditional probability of (12a) equals p for all possible spaces $A(\mathbf{y})$ then the unconditional probability is also p.

If the fiducial function is integrated over regions other than those of the type (12) but such that the integral equals p, then there is no guarantee that the statement arrived at by replacing $\mathbf{y}^0$ by $\mathbf{y}$ and $(\mathbf{b}, s)$ by $(\boldsymbol{\beta}, \sigma)$ as in (12) and (12a), is satisfied with probability p. Pitman (1957) discusses the shortcomings of the Behrens–Fisher test in relation to 'proper regions'.

SIMULTANEOUS AND INDIVIDUAL CONFIDENCE SETS FOR THE PARAMETERS

For a given space $A(\mathbf{y}^0)$ the integration region defined by (13)

$$\left.\begin{aligned} s &\subset S[A(y^0)], \\ b_\nu &\subset B_\nu[A(y^0)] \quad (\nu = 1, \ldots, r) \end{aligned}\right\} \tag{13}$$

is proper since it may be written in the form (12) by defining

$$P \cap A(\mathbf{y}^0) \equiv \left\{\frac{\mathbf{y}^0 - \mathbf{Xb}}{s} : s \subset S[A(y^0)],\ b_\nu \subset B_\nu[A(\mathbf{y}^0)],\ \nu = 1, \ldots, r\right\}.$$

Hence by integrating the fiducial function corresponding to $\mathbf{y}^0$ over the region (13) we obtain the conditional probability that

$$\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma} \subset P \cap A(\mathbf{y}^0), \tag{14}$$

given $\mathbf{y} \subset A(\mathbf{y}^0)$. Since for $\mathbf{y} \subset A(\mathbf{y}^0)$ the vectors $\mathbf{y}$ may be written $(\mathbf{y}^0 - \mathbf{Xb})/s$ (see formula (2)), formula (14) reduces to

$$\frac{\mathbf{y}^0}{\sigma s} - \mathbf{X}\left(\frac{\mathbf{b} + s\boldsymbol{\beta}}{\sigma s}\right) \subset \left\{\frac{\mathbf{y}^0}{s} - \frac{\mathbf{Xb}}{s} : s \subset S[A(\mathbf{y}^0)],\ b_\nu \subset B_\nu[A(\mathbf{y}^0)]\right\} \tag{15}$$

which is equivalent to the set of relations

$$\left.\begin{aligned} b_\nu + \beta_\nu s &\subset B_\nu[A(\mathbf{y}^0)] \quad (\nu = 1, \ldots, r), \\ \sigma s &\subset S[A(\mathbf{y}^0)]. \end{aligned}\right\} \tag{16}$$

The relations (16) define simultaneous confidence sets for the parameters and are satisfied with conditional probability p.

Individual confidence relations for the parameter B_i say are obtained by taking $S[A(\mathbf{y}^0)]$ as the range $(0, \infty)$ and $B_\nu[A(\mathbf{y}^0)]$ as the range $(-\infty, \infty)$ for all $\nu \neq i$. The set $B_i[A(\mathbf{y}^0)]$ must be defined to give the integral of the fiducial function over the region

$$b_i \subset B_i[A(\mathbf{y}^0)]$$

the required value say p. Individual confidence sets for σ are found in a similar manner. If for each possible space $A(\mathbf{y})$ confidence relations are constructed which hold with conditional probability p then the unconditional probability is also p.

THE ESTIMATOR PROPERTY

Any two sets of parameter values $(\boldsymbol{\beta}_1, \sigma_1)$ and $(\boldsymbol{\beta}_2, \sigma_2)$ are related by a transformation of the type

$$\left.\begin{aligned}\boldsymbol{\beta}_2 &= \lambda\boldsymbol{\beta}_1 + \lambda\boldsymbol{\Delta},\\ \sigma_2 &= \lambda\sigma_1\end{aligned}\right\} \tag{17}$$

for some $\lambda > 0$ and $\boldsymbol{\Delta}$. If in $\qquad \mathbf{y}_1 = \mathbf{X}\boldsymbol{\beta}_1 + \sigma_1\mathbf{u}$

the parameter values $(\boldsymbol{\beta}_1, \sigma_1)$ are replaced by $(\boldsymbol{\beta}_2, \sigma_2)$, then the distribution of $\mathbf{y}_2$ may be obtained from that of $\mathbf{y}_1$ by the transformation

$$\mathbf{y}_2 = \lambda(\mathbf{y}_1 + \mathbf{X}\boldsymbol{\Delta}). \tag{18}$$

A statistic is said to have the estimator property when it mimics the parameter it is to estimate, in that, just as the change (17) in parameters leads to the transformation (18), estimates for the parameters corresponding to $\mathbf{y}_1$ and $\mathbf{y}_2$ are related by

$$\left.\begin{aligned}\hat{\boldsymbol{\beta}}_2 &= \lambda\hat{\boldsymbol{\beta}}_1 + \lambda\boldsymbol{\Delta},\\ \hat{\sigma}_2 &= \lambda\hat{\sigma}_1.\end{aligned}\right\} \tag{19}$$

the analogue of (17).† More explicitly (19) may be written

$$\begin{aligned}\hat{\beta}_\nu(\lambda[\mathbf{y} + \mathbf{X}\boldsymbol{\Delta}]) &= \lambda\hat{\beta}_\nu(\mathbf{y}) + \lambda\Delta_\nu, \quad (\nu = 1, \ldots, r),\\ \hat{\sigma}(\lambda[\mathbf{y} + \mathbf{X}\boldsymbol{\Delta}]) &= \lambda\hat{\sigma}(\mathbf{y}).\end{aligned} \tag{20}$$

for all $\lambda > 0$ and all $\boldsymbol{\Delta}$.

The definition of the estimator property is due to Pitman (1938). It is introduced here in a slightly different manner and for a more general situation. As in the case of location and scale parameters the concept defines a class of statistics readily dealt with in terms of fiducial functions. Commonly used statistics such as the least-squares estimators and the maximum-likelihood estimators have the estimator property. See examples 1 and 2 below.

Example 1. The least-squares estimators $\hat{\boldsymbol{\beta}}^\square$ and the associated estimator $\hat{\sigma}^\square$ have the estimator property.‡

The least-squares estimators $\hat{\boldsymbol{\beta}}^\square$ for the regression parameters $\boldsymbol{\beta}$ are given by

$$\hat{\boldsymbol{\beta}}^\square = (\mathbf{X}' \vdots \mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Replacing $\mathbf{y}$ by $\lambda(\mathbf{y} + \mathbf{X}\boldsymbol{\Delta})$ gives

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\lambda\mathbf{y} + \lambda\mathbf{X}\boldsymbol{\Delta}) = \lambda\hat{\boldsymbol{\beta}}^\square + \lambda\boldsymbol{\Delta}$$

as required by (20). Further, supposing all u_i have unit variance, the familiar estimator $\hat{\sigma}^\square$ for the error-scale parameter is

$$\hat{\sigma}^\square = \sqrt{\left(\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^\square)'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^\square)}{n - r}\right)}.$$

Replacing $\mathbf{y}$ by $\lambda(\mathbf{y} + \mathbf{X}\boldsymbol{\Delta})$ gives $\lambda\hat{\sigma}^\square$ as required by (20).

† All estimators in this paper will carry the circumflex sign $\hat{}$.
‡ All least-squares estimators are labelled with a square $\square$ superscript.

Example 2. The maximum-likelihood estimators for $\boldsymbol{\beta}$ and σ have the estimator property. The maximum-likelihood estimators corresponding to $\mathbf{y}$ simultaneously satisfy

$$\left.\begin{aligned} \frac{\partial}{\partial\beta_\nu} F\left(\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) &= 0 \quad (\nu = 1, \ldots, r), \\ \frac{\partial}{\partial\sigma} F\left(\frac{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) &= 0. \end{aligned}\right\} \tag{21}$$

The maximum-likelihood estimators corresponding to $\lambda\mathbf{y}+\lambda\mathbf{X}\boldsymbol{\Delta}$ simultaneously satisfy

$$\left.\begin{aligned} \frac{\partial}{\partial\beta_\nu} F\left(\frac{\mathbf{y}-\mathbf{X}\{(\boldsymbol{\beta}/\lambda)-\boldsymbol{\Delta}\}}{\sigma/\lambda}\right) &= 0 \quad (\nu = 1, \ldots, r), \\ \frac{\partial}{\partial\sigma} F\left(\frac{\mathbf{y}-\mathbf{X}\{(\boldsymbol{\beta}/\lambda)-\boldsymbol{\Delta}\}}{\sigma/\lambda}\right) &= 0. \end{aligned}\right\} \tag{22}$$

The equations (22) may also be written as

$$\left.\begin{aligned} \frac{\partial}{\lambda\,\partial\{(\beta_\nu/\lambda)-\Delta_\nu\}} F\left(\frac{\mathbf{y}-\mathbf{X}\{(\boldsymbol{\beta}/\lambda)-\boldsymbol{\Delta}\}}{\sigma/\lambda}\right) &= 0 \quad (\nu = 1, \ldots, r), \\ \frac{\partial}{\lambda\,\partial(\sigma/\lambda)} F\left(\frac{\mathbf{y}-\mathbf{X}\{(\boldsymbol{\beta}/\lambda)-\boldsymbol{\Delta}\}}{\sigma/\lambda}\right) &= 0. \end{aligned}\right\} \tag{23}$$

The solutions of (21) for $\beta_1, \ldots, \beta_r$ and σ are related to the solutions of (23) by (20) as required.

Example 3. Fixed quantile points of the marginal fiducial functions corresponding to vectors $\mathbf{y} \subset A(\mathbf{y}^0)$ have the estimator property.

Any vector $\mathbf{y} \subset A(\mathbf{y}^0)$ may be written as $\lambda(\mathbf{y}^0+\mathbf{X}\boldsymbol{\Delta})$ for some $\lambda > 0$ and $\boldsymbol{\Delta}$. The fiducial function corresponding to $\mathbf{y}$ is

$$K(\mathbf{w})\, F\left(\frac{\mathbf{y}^0-\mathbf{X}\{(\mathbf{b}/\lambda)-\boldsymbol{\Delta}\}}{s/\lambda}\right)\frac{1}{(s/\lambda)^{n+1}}.$$

It follows that the marginal quantile points transform according to (20), i.e. they have the estimator property.

Remark 1. If two statistics for the same parameter both have the estimator property for all $\mathbf{y} \subset A(\mathbf{y}^0)$ then their difference does not change sign. This follows from (20) by subtracting two such estimators.

Remark 2. If for different spaces $A(\mathbf{y})$ a statistic is defined by a different marginal quantile point it will still have the estimator property. The estimator property only relates estimators corresponding to different vectors $\mathbf{y}$ within the same space $A(\mathbf{y})$.

Closest estimators

If for all estimators $\hat{\theta}$ in a class, the estimator $\hat{\theta}^c$ is such that

$$|\hat{\theta}^c-\theta| \leqslant |\hat{\theta}-\theta|$$

with probability $\geqslant \frac{1}{2}$, $\hat{\theta}^c$ will be called a closest estimator for θ in the class.†

Proposition. The median $\hat{\theta}^c$ of the marginal fiducial function for that parameter is a closest estimator.

† All closest estimators in this paper will be labelled with a superscript c.

If $\hat{\theta}^c$ stands for the median of the marginal fiducial function for θ, then the random interval $(-\infty, \hat{\theta}^c)$ has probability 0·5 of containing the true parameter value θ. Estimators are either smaller or larger than $\hat{\theta}^c$ for all $\mathbf{y} \subset A(\mathbf{y}^0)$ (*Remark* 1 in the previous section). Those which are larger will be more distant from the true parameter value θ at least when the interval $(-\infty, \hat{\theta}^c)$ contains θ, i.e. in at least 50 % of cases. The reasoning for estimators smaller than $\hat{\theta}^c$ is the same when instead the interval $(\hat{\theta}^c, \infty)$ is used.

Example 4. Closest estimators of the regression and error-scale parameters, when the errors are normal and independently distributed with zero means and unknown equal variances.

The joint distribution of $y_1, \ldots, y_n$ readily reduces to

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2}\left[\frac{(n-r)\,(\hat{\sigma}^\square)^2}{\sigma^2} + \left(\frac{\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^\square}{\sigma}\right)' \mathbf{X}'\mathbf{X}\left(\frac{\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^\square}{\sigma}\right)\right]\right\} dy_1, \ldots, dy_n,$$

where $\hat{\boldsymbol{\beta}}^\square$ is the vector of least-squares estimators for $\boldsymbol{\beta}$ and $\hat{\sigma}^\square$ is the associated estimator for σ (see example 1 of the preceding section).

The fiducial function for the parameters $\beta_1, \ldots, \beta_r, \sigma$ is

$$K(\mathbf{w})\frac{1}{s^{n+1}}\exp\left\{-\frac{(n-r)\,(\hat{\sigma}^\square)^2}{2s^2} - \frac{(\hat{\boldsymbol{\beta}}^\square - \mathbf{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}^\square - \mathbf{b})}{2s^2}\right\}.$$

The closest estimator $\hat{\beta}_\nu^c$ of β_ν equals $\hat{\beta}_\nu^\square$, the mean of the symmetrical marginal fiducial function for β_ν ($\nu = 1, \ldots, r$). Integrating out $\beta_1, \ldots, \beta_r$ from the fiducial function gives the marginal fiducial function for σ and its median $\hat{\sigma}^c$ is defined by

$$\int_{s=0}^{\hat{\sigma}^c} K\frac{1}{s^{n+1-r}}\exp\left\{-\frac{(n-r)\,(\hat{\sigma}^\square)^2}{2s^2}\right\} ds = \tfrac{1}{2}.$$

The transformation $\xi = \frac{1}{2}(n-r)\,(\hat{\sigma}^\square)^2/s^2$ gives

$$\int_{\xi=\frac{1}{2}(n-r)\,(\hat{\sigma}^\square/\hat{\sigma}^c)^2}^{\infty} \xi^{\frac{1}{2}(n-r-2)}\frac{e^{-\xi}}{\Gamma\{\frac{1}{2}(n-r)\}} d\xi = \tfrac{1}{2}.$$

The lower limit of integration is the median of a gamma distribution $\Gamma\{\frac{1}{2}(n-r)\}$. Pitman (1937) has shown that the median of a $\Gamma\{\frac{1}{2}(n-r)\}$ distribution is approximately $\{\frac{1}{2}(n-r)-\frac{1}{3}\}$. It follows

$$\left(\frac{n-r}{2}\right)\left(\frac{\hat{\sigma}^\square}{\hat{\sigma}^c}\right)^2 \approx \left(\frac{n-r}{2}\right) - \frac{1}{3}.$$

The closest estimator $\hat{\sigma}^c$ is slightly larger than $\hat{\sigma}^\square$.

Least-mean-square-error estimators

The estimator $\hat{\beta}^m$ which makes the expected value $E(\hat{\beta}_\nu^m - \beta_\nu)^2$ of $(\hat{\beta}_\nu^m - \beta_\nu)^2$ a minimum will be called the least-mean-square-error estimator.† For $\mathbf{y} \subset A(\mathbf{y}^0)$ the transformation used in deriving the fiducial function (10) may be written as (6). With the aid of (6) and the estimator property (20), $(\hat{\beta}_\nu(y) - \beta_\nu)^2$ reduces to

$$\frac{\sigma^2}{s^2}(\hat{\beta}_\nu(\mathbf{y}^0) - b_\nu)^2.$$

† Least-mean-square-error estimators will be labelled with a superscript m.

Hence $E(\hat{\beta}_\nu - \beta_\nu)^2$ may be written as

$$\sigma^2 \iint\cdots\int_{w_1 \dots w_{n-r-1}} \psi(\mathbf{w}) \left[\iiint \left(\frac{\hat{\beta}_\nu - b_\nu}{s} \right)^2 \frac{1}{s^{n+1}} F\left(\frac{\mathbf{y}^0 - \mathbf{X}\mathbf{b}}{s} \right) db_1 \dots db_r ds \right] dw_1 \dots dw_{n-r-1}.$$

This integral will be a minimum if the part in the square brackets, i.e.

$$E_f\left(\frac{\hat{\beta}_\nu - b_\nu}{s} \right)^2$$

is a minimum. (f is mnemonic for fiducial.) Thus $\hat{\beta}_\nu^m$ is the solution of

$$\frac{\partial}{\partial \hat{\beta}_\nu} E_f\left(\frac{\hat{\beta}_\nu - b_\nu}{s} \right) = 0,$$

this gives

$$\hat{\beta}_\nu^m = \frac{E_f(b_\nu/s^2)}{E_f(1/s^2)}.$$

Similarly the least-mean-square-error estimator $\hat{\sigma}^m$ of σ is the one which makes

$$E(\hat{\sigma}^m - \sigma)^2$$

a minimum. As before it follows $E_f\{(\hat{\sigma}^m - s)/s\}^2$ must be a minimum. By putting

$$\frac{\partial}{\partial \hat{\sigma}} E_f\left(\frac{\hat{\sigma} - s}{s} \right)^2 = 0$$

it follows that

$$\hat{\sigma}^m = \frac{E_f(1/s)}{E_f(1/s^2)}.$$

Example 5. For the case of normal independently distributed errors it turns out

$$\hat{\beta}_\nu^m = \hat{\beta}_\nu^\square \quad (\nu = 1, \dots, r),$$

$$\hat{\sigma}^m = \hat{\sigma}^\square \sqrt{\frac{n-r}{2}} \frac{\Gamma\{\frac{1}{2}(n-r+1)\}}{\Gamma\{\frac{1}{2}(n-r+2)\}} \to \hat{\sigma}^\square \quad \text{as} \quad n \to \infty.$$

Thanks are due to Dr E. A. Cornish, Dr G. H. Jowett and a referee for helpful comments on the manuscript.

REFERENCES

PITMAN, E. J. G. (1937). The 'closest' estimators of statistical parameters. *Proc. Camb. Phil. Soc.* **33**, 212–22.

PITMAN, E. J. G. (1938). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika*, **30**, 391–421.

PITMAN, E. J. G. (1957). Statistics and science. *J. Amer. Stat. Assn.* **52**, 322–30.

Biometrika (1961), **48**, 1 *and* 2, *p.* 133
Printed in Great Britain

Latent vectors of random symmetric matrices†

BY C. L. MALLOWS
University College, London‡

SUMMARY

Matrices of random variables that are symmetric but not necessarily positive definite can arise in various ways. Assuming normality and a certain 'invariant' covariance structure, or alternatively assuming a Wishart structure, we derive exact likelihood ratio tests of certain hypotheses on the latent vectors. In the former case, these lead to exact confidence regions for the unknown vectors; approximations to these regions are given. The approximate validity of the regions in non-normal and non-invariant cases (for example, in the case of rounding-off errors) is investigated. An application to the analysis of a quadratic response surface is described.

1. INTRODUCTION

We consider square symmetric matrices, not necessarily positive definite, whose elements are random variables. Such matrices can arise in several ways, for example:

(i) in numerical work where the elements of a matrix are computed only to a certain number of decimals, the rounding-off errors may be regarded as being independent uniform random variables;

(ii) more generally, the cumulative effect of rounding-off errors at various stages of a computation may lead to errors in a final matrix that are individually 'between' rectangular and normal, and which in general are not independent;

(iii) in the fitting of multifactor quadratic response surfaces one obtains a matrix of second-order regression coefficients, the elements of which are linear in the observations; these elements may be regarded as being normally distributed, and their covariance structure depends on the design of the experiment;

(iv) many multivariate problems lead to analyses in terms of sample covariance matrices, which are necessarily positive definite, and may be assumed to have the Wishart distribution.

In § 2 we state various sets of precise assumptions regarding the distribution of the elements of the random matrix. Various hypotheses regarding the latent vectors are given in § 3, and the corresponding test statistics are derived in § 4. In § 5, the distributions of the normal-theory statistics are investigated in non-normal situations. Confidence regions are derived in § 6, expressions in statistical differentials are used to derive some refined approximations in § 7, and § 8 treats a numerical example from two points of view.

2. ASSUMPTIONS

We shall assume throughout that the observed $(k \times k)$ matrix $\mathbf{X}$ has the structure

$$\mathbf{X} = \mathbf{\Xi} + \mathbf{Z},$$

† This paper is based partly on Technical Report No. 35 of the Statistical Techniques Research Group of Princeton University, which was written with the support of the Office of Ordnance Research, U.S. Army while the author was on leave of absence from University College, London.

‡ Now at Columbia University and Bell Telephone Laboratories.

where $\boldsymbol{\Xi}$ is a fixed symmetric matrix, and $\mathbf{Z}$ is a symmetric matrix of random variables with zero means. Thus

$$\mathscr{E}(\mathbf{Z}) = \mathbf{O}, \quad \mathscr{E}(\mathbf{X}) = \boldsymbol{\Xi}.$$

We shall write

$$\boldsymbol{\Xi} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T, \quad \mathbf{X} = \mathbf{G}\mathbf{L}\mathbf{G}^T, \quad \mathbf{H} = \mathbf{G}^T\boldsymbol{\Gamma},$$

where $\boldsymbol{\Gamma}$ and $\mathbf{G}$ are orthogonal matrices whose columns form normalized latent vectors of $\boldsymbol{\Xi}$ and $\mathbf{X}$, respectively, and where $\boldsymbol{\Lambda}$ and $\mathbf{L}$ are diagonal

$$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k), \quad \mathbf{L} = \operatorname{diag}(l_1, l_2, \ldots, l_k).$$

When necessary for definiteness, we shall assume that

$$\Gamma_{i1} \geqslant 0, \quad G_{i1} \geqslant 0, \quad \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k, \quad l_1 \geqslant l_2 \geqslant \ldots \geqslant l_k. \tag{2·1}$$

Various assumptions may be appropriate regarding the covariance structure and distributional properties of the elements of $\mathbf{Z}$. Here we shall assume that the covariance structure is known up to a constant factor, so that if we write

$$\operatorname{cov}(Z_{ij}, Z_{lm}) = V_{ij,lm}\sigma^2$$

then all the V's are known. If the Z's are rounding-off errors, then σ^2 will be known also; otherwise we may require an independent estimate of σ^2. We shall often assume that the elements of $\mathbf{Z}$ are either (*a*) multinormal or (*b*) uniform.

The following sets of more detailed assumptions will lead to especially simple results.

(i) *The invariant covariance structure*

It is shown in Mallows (1959) (referred to hereafter as T.R. 35) that a necessary and sufficient condition for the covariance structure of a random symmetric matrix to be invariant under all transformations of the form $\mathbf{Z}' = \mathbf{P}\mathbf{Z}\mathbf{P}^T$ (with $\mathbf{P}$ orthogonal) is

$$\operatorname{cov}(Z_{ij}, Z_{lm}) = \{v(\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl}) + c\delta_{ij}\delta_{lm}\}\sigma^2 \tag{2·2}$$

for some σ^2, v, c. Thus in particular

$$\operatorname{var}(Z_{ij}) = v\sigma^2 \quad (i \neq j), \qquad \operatorname{cov}(Z_{ii}, Z_{jj}) = c\sigma^2 \quad (i \neq j), \qquad \operatorname{var}(Z_{ii}) = (2v + c)\sigma^2,$$

and all other covariances vanish.

For $c \geqslant 0$, such a matrix $\mathbf{Z}$ can be represented as

$$\mathbf{Z} = w\mathbf{I}_k + \mathbf{W} + \mathbf{W}^T, \tag{2·3}$$

where w and all the elements of $\mathbf{W}$ are uncorrelated variables, with

$$\operatorname{var}(w) = c\sigma^2, \quad \operatorname{var}(W_{ij}) = \tfrac{1}{2}v\sigma^2.$$

This structure arises naturally in a certain response-surface problem (see § 8). Here, the elements of $\mathbf{X}$ are linear in the observations, and it is reasonable to assume that they are normal; neither $\boldsymbol{\Xi}$ nor $\mathbf{X}$ is necessarily positive definite. v and c are functions of the experimental design, and depend on the moments of the design up to the fourth order.

If it is assumed that the observations are not normal, the means and covariances of the elements of $\mathbf{X}$ are unaffected; their fourth moments are effected by the moments of the design up to the *eighth* order. We have not attempted to investigate this effect.

However, there is another reasonably simple non-normal modification which has an application to a case involving rounding-off errors. It is obtained by assuming that in the representation (2·3) we have $c = 0$, $v = 1$, and the elements of $\mathbf{W}$ are independent with

$$\text{var}\,(W_{ij}) = \tfrac{1}{2}\sigma^2, \quad \mu_3(W_{ij}) = 0, \quad \mu_4(W_{ij}) = (\tfrac{1}{2}\sigma^2)^2(3+\kappa_4). \tag{2·4}$$

This leads to

$$\left.\begin{aligned} \text{var}\,(Z_{ij}) &= \sigma^2 \quad (i \neq j), & \mu_4(Z_{ij}) &= \sigma^4(3+\tfrac{1}{2}\kappa_4) \quad (i \neq j), \\ &= 2\sigma^2 \quad (i=j), & &= 4\sigma^4(3+\kappa_4) \quad (i=j), \\ \mu_{22}(Z_{ij}, Z_{lm}) &= \text{var}\,(Z_{ij})\,\text{var}\,(Z_{lm}) \quad (i,j) \neq (l,m) \text{ or } (m,l). \end{aligned}\right\} \tag{2·5}$$

with all odd moments (up to the fourth order) zero. If the elements of $\mathbf{W}$ are independent rounding-off errors, then $\kappa_4 = -\frac{6}{5}$.

(ii) *The Wishart covariance structure*

Here the covariance structure of the elements of $\mathbf{X}$ is the same as that of a matrix of sample variances and covariances in a sample of $n+1$ from a k-variate population with covariance matrix $\boldsymbol{\Xi}$, and zero third and fourth cumulants. We have

$$\left.\begin{aligned} \mathscr{E}(X_{ij}) &= \Xi_{ij}, \\ \text{cov}\,(X_{ij}, X_{lm}) &= n^{-1}(\Xi_{il}\Xi_{jm} + \Xi_{im}\Xi_{jl}). \end{aligned}\right\} \tag{2·6}$$

If $\mathbf{C}$ is a matrix such that $\mathbf{C}\boldsymbol{\Xi}\mathbf{C}^T = \mathbf{I}_k$, then the matrix

$$\mathbf{Y} = \mathbf{C}\mathbf{X}\mathbf{C}^T - \mathbf{I}_k$$

has zero expectation and has the 'Unit Wishart' covariance structure, namely

$$\text{cov}\,(Y_{ij}, Y_{lm}) = n^{-1}(\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl}).$$

This is a special case of the invariant structure (2·2).

If the underlying k-variate population is multinormal, we shall as usual say that $\mathbf{X}$ has the Wishart distribution with n d.f., based on the covariance matrix $\boldsymbol{\Xi}$; in this case $\mathbf{X}$ must be positive definite. If the higher cumulants of the k-variate population are not zero, $\mathscr{E}(X_{ij})$ is unchanged, but now

$$\text{cov}\,(X_{ij}, X_{lm}) = n^{-1}(\Xi_{il}\Xi_{jm} + \Xi_{im}\Xi_{jl}) + (n+1)^{-1}\kappa_{ijlm}.$$

(iii) *The rounding-off structure*

Here we assume that the elements of $\mathbf{X}$ are rounded off directly; so that the diagonal and super-diagonal elements of $\mathbf{Z}$ are independent uniform variables, with

$$\text{var}\,(Z_{ij}) = \sigma^2, \quad \mu_3(Z_{ij}) = 0, \quad \mu_4(Z_{ij}) = \sigma^4(3+\kappa_4) \tag{2·7}$$

where $\kappa_4 = -\frac{6}{5}$.

3. Hypotheses considered

Consider the following hypothesis regarding latent vectors (l.v.'s) of $\boldsymbol{\Xi}$

$H_p(\boldsymbol{\Gamma}_p)$: A specified set of $p(\leqslant k)$ orthogonal unit vectors $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_p$ forming the columns of a $(k \times p)$ matrix $\boldsymbol{\Gamma}_p$ are l.v.'s of $\boldsymbol{\Xi}$, i.e.

$$\boldsymbol{\Xi}\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i\lambda_i \quad \text{for some } \lambda_i \ (i = 1, \ldots, p),$$

or equivalently

$$(\mathbf{I}_k - \boldsymbol{\Gamma}_p\boldsymbol{\Gamma}_p^T)\boldsymbol{\Xi}\boldsymbol{\Gamma}_p = \mathbf{0}. \tag{3·1}$$

This hypothesis imposes $d_p = p(k-1) - \frac{1}{2}p(p-1)$

independent linear constraints on the elements of Ξ; it is therefore a linear hypothesis of this order. $H_p(\Gamma_p)$ is the intersection of the hypotheses $H_1(\gamma_1), \ldots, H_1(\gamma_p)$; $H_k(\Gamma_k)$ asserts that $\Gamma_k^T \Xi \Gamma_k$ is diagonal.

Notice that the order of the columns of Γ_p is not important; we are not here imposing the restriction that the corresponding latent roots (l.r.'s) are ordered.

Although $H_p(\Gamma_p)$ is a linear hypothesis, the elements of Γ_p are not linear parameters. The matrix Ξ can be specified by any set of $d_\Omega = \frac{1}{2}k(k+1)$ functionally independent elements, for example, by the elements in and above the diagonal. Thus Ξ can be represented by a point ξ in a d_Ω-dimensional space Ω. $H_p(\Gamma_p)$ asserts that ξ lies in a linear subspace of $d_\omega = d_\Omega - d_p$ dimensions. However (if $k \geqslant 3$), for any one pair (i,j), if Γ_{ij} is held fixed, the locus of points ξ corresponding to matrices Ξ satisfying (3·1) for some values of the remaining elements of Γ_p, is *not* a linear subspace. This implies that the least-squares estimates of the elements of Γ_p are not linear in the elements of $\mathbf{X}$; and there do not exist d_p linear functions of the elements of $\mathbf{X}$ which are sufficient to determine these estimates. This does not affect the possibility of constructing exact confidence regions for the elements of Γ_p (compare Beale (1960), especially pp. 44–5).

We can consider also the linear hypothesis $H_p(\Gamma_p; \Lambda_p)$ which specifies p l.v.'s of Ξ and also the associated l.r.'s; and the linear conditional hypothesis $H_p(\Lambda_p | \Gamma_p)$ which specifies the l.r.'s conditionally on the l.v.'s. These hypotheses impose respectively $p + d_p$ and p constraints on the elements of Ξ. $H_k(\Gamma_k; \Lambda_k)$ specifies Ξ completely.

There are many possibilities in the construction of systems of nested linear hypotheses; to preserve the linearity of the set-up, the hypotheses introduced at each stage must never specify a l.r. before the corresponding l.v.

4. Derivation of test criteria

4.1. *Normal invariant case.* We derive the likelihood ratio tests of the hypotheses described in the previous section, relative to the alternative hypothesis that Ξ is completely arbitrary. When the elements of $\mathbf{Z}$ are multinormal with zero means and with the invariant covariance structure (2·2), the log likelihood is

$$L(\Xi) = \text{const.} - \tfrac{1}{2}k(k+1)\ln\sigma^2 - \tfrac{1}{2}S(\Xi), \tag{4·1}$$

where $$S(\Xi) = \frac{1}{2v\sigma^2}\left(\operatorname{tr}(\mathbf{X}-\Xi)^2 - \frac{c}{2v+kc}[\operatorname{tr}(\mathbf{X}-\Xi)]^2\right).$$

Treating σ^2 as known for the present, the unrestricted maximum is obtained by setting $\hat{\Xi} = \mathbf{X}$, and we have $$\min S(\Xi) = S(\mathbf{X}) = 0.$$

In maximizing $L(\Xi)$ subject to the $p + d_p$ constraints on Ξ imposed by $H_p(\Gamma_p; \Lambda_p)$, it would seem that we should include terms containing just this number of Lagrange multipliers. However, we shall proceed by including a full $(k \times p)$ matrix θ of multipliers, one for each of the (non-independent) constraints in

$$\Xi\Gamma_p = \Gamma_p\Lambda_p.$$

The resulting equations will not be sufficient to determine all the Lagrange multipliers uniquely, but will be sufficient to determine the maximum of $L(\Xi)$. Thus we minimize

$$S(\Xi) + \operatorname{tr}(\theta^T(\Xi\Gamma_p - \Gamma_p\Lambda_p))$$

obtaining after some algebra the test statistic

$$\begin{aligned}\psi_p(\boldsymbol{\Gamma}_p;\boldsymbol{\Lambda}_p) &= \min_{H_p} S(\boldsymbol{\Xi})\\ &= \frac{1}{2v\sigma^2}\{2\,\mathrm{tr}\,(\boldsymbol{\Gamma}_p^T\mathbf{X}(\mathbf{I}_k-\boldsymbol{\Gamma}_p\boldsymbol{\Gamma}_p^T)\mathbf{X}\boldsymbol{\Gamma}_p)+\mathrm{tr}\,(\boldsymbol{\Gamma}_p^T\mathbf{X}\boldsymbol{\Gamma}_p-\hat{\boldsymbol{\Lambda}}_p)^2\}\\ &\quad+\frac{1}{2v\sigma^2}\left\{\mathrm{tr}\,(\hat{\boldsymbol{\Lambda}}_p-\boldsymbol{\Lambda}_p)^2-\frac{c}{2v+pc}[\mathrm{tr}\,(\hat{\boldsymbol{\Lambda}}_p-\boldsymbol{\Lambda}_p)]^2\right\} \qquad (4{\cdot}2)\\ &= \psi_p(\boldsymbol{\Gamma}_p)+\psi_p(\boldsymbol{\Lambda}_p|\boldsymbol{\Gamma}_p),\end{aligned}$$

where $\hat{\boldsymbol{\Lambda}}_p$ $(p\times p)$ contains the diagonal elements of $\boldsymbol{\Gamma}_p^T\mathbf{X}\boldsymbol{\Gamma}_p$, with zeros elsewhere. (Van der Vaart (1958) remarks that $\hat{\boldsymbol{\Lambda}}_p$ is an unbiased estimator of $\boldsymbol{\Lambda}_p$, while the elements of $\mathbf{L}$ do not give unbiased estimates.) $\psi_p(\boldsymbol{\Gamma}_p;\boldsymbol{\Lambda}_p)$ is distributed as χ^2 with $p+d_p$ d.f., and with non-centrality obtained from (4·2) by replacing $\mathbf{X}$ by $\boldsymbol{\Xi}$ throughout. The non-centrality is zero if $H_p(\boldsymbol{\Gamma}_p;\boldsymbol{\Lambda}_p)$ is true.

For $H_p(\boldsymbol{\Gamma}_p)$, we now minimize (4·2) with respect to $\boldsymbol{\Lambda}_p$. Hence the statistic $\psi_p(\boldsymbol{\Gamma}_p)$ for testing $H_p(\boldsymbol{\Gamma}_p)$ is given by the first two terms of (4·2), and $\psi_p(\boldsymbol{\Lambda}_p|\boldsymbol{\Gamma}_p)$ by the last two; the degrees of freedom are d_p and p, respectively.

Putting $p=1$, and writing $\boldsymbol{\gamma}$, λ for $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Lambda}_1$ we find

$$\begin{aligned}\psi_1(\boldsymbol{\gamma};\lambda) &= \frac{1}{v\sigma^2}\{\boldsymbol{\gamma}^T\mathbf{X}^2\boldsymbol{\gamma}-(\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma})^2\}+\frac{1}{(2v+c)\,\sigma^2}(\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}-\lambda)^2 \qquad (4{\cdot}3)\\ &= \psi_1(\boldsymbol{\gamma})+\psi_1(\lambda|\boldsymbol{\gamma}).\end{aligned}$$

Putting $p=k$, the first term in (4·2) vanishes, and $\psi_k(\boldsymbol{\Gamma}_k)$ can be written in the forms

$$\begin{aligned}\psi_k(\boldsymbol{\Gamma}_k) &= \frac{1}{v\sigma^2}\sum_{i<j}[(\boldsymbol{\Gamma}_k^T\mathbf{X}\boldsymbol{\Gamma}_k)_{ij}]^2\\ &= \frac{1}{2v\sigma^2}\{\sum_{ij}\mathbf{X}_{ij}^2-\sum_i[(\boldsymbol{\Gamma}_k^T\mathbf{X}\boldsymbol{\Gamma}_k)_{ii}]^2\} \qquad (4{\cdot}4)\\ &= \tfrac{1}{2}\sum_i\psi_1(\boldsymbol{\gamma}_i)\end{aligned}$$

where $\psi_1(\boldsymbol{\gamma}_i)$ is $\psi_1(\boldsymbol{\gamma})$ computed for the ith column of $\boldsymbol{\Gamma}_k$. Also

$$\psi_k(\boldsymbol{\Lambda}_k|\boldsymbol{\Gamma}_k) = \mathbf{1}_k^T(\hat{\boldsymbol{\Lambda}}_k-\boldsymbol{\Lambda}_k)\,\mathbf{V}^{-1}(\hat{\boldsymbol{\Lambda}}_k-\boldsymbol{\Lambda}_k)\,\mathbf{1}_k, \qquad (4{\cdot}5)$$

where $\mathbf{1}_k$ is a $(k\times 1)$ column vector of ones, so that $\boldsymbol{\Lambda}_k\mathbf{1}_k$ is a column vector formed from the diagonal elements of $\boldsymbol{\Lambda}_k$; and $\mathbf{V}$ is the covariance matrix of the diagonal elements of $\mathbf{X}$, i.e.

$$\begin{aligned}\mathbf{V} &= 2v\sigma^2\mathbf{I}_k+c\sigma^2\mathbf{1}_k\mathbf{1}_k^T,\\ \mathbf{V}^{-1} &= (2v\sigma^2)^{-1}(\mathbf{I}_k-c(2v+kc)^{-1}\mathbf{1}_k\mathbf{1}_k^T).\end{aligned}$$

If σ^2 is unknown but an independent estimate s^2 is available which is distributed as $\sigma^2\chi_f^2/f$, the obvious modifications have to be made to the above derivations. The likelihood ratio criteria are now monotone functions of the statistics, obtained from the ψ's given above by replacing σ^2 by s^2 throughout; the distribution of the modified $\psi_p(\boldsymbol{\Gamma}_p)$, for example, becomes that of F with d_p and f degrees of freedom. When dealing with a conditional hypothesis, an estimate of σ^2 becomes available from $\mathbf{X}$ itself. Thus to test the hypothesis $H_1(\lambda|\boldsymbol{\gamma})$ we may compare

$$\frac{(k-1)\,\psi_1(\lambda|\boldsymbol{\gamma})}{\psi_1(\boldsymbol{\gamma})} = \frac{(k-1)v}{2v+c}\,\frac{(\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}-\lambda)^2}{\boldsymbol{\gamma}^T\mathbf{X}^2\boldsymbol{\gamma}-(\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma})^2},$$

with the F distribution on 1, $k-1$ d.f. The procedure can be presented as an analysis of variance.

4.2. *General nested hypotheses.* Suppose the hypothesis H specifies $p = p_1 + p_2$ l.v.'s and p_1 of the corresponding l.r.'s. We now need to minimize $S(\boldsymbol{\Xi})$ subject to the constraints

$$\boldsymbol{\Xi}\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1, \quad \boldsymbol{\Xi}\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}_2\boldsymbol{\Lambda}_2, \tag{4·6}$$

where $\boldsymbol{\Gamma}_1(k \times p_1)$, $\boldsymbol{\Gamma}_2(k \times p_2)$ and $\boldsymbol{\Lambda}_1((p_1 \times p_1)$ diagonal) are specified by the hypothesis, but $\boldsymbol{\Lambda}_2((p_2 \times p_2)$ diagonal) is arbitrary. Minimizing first with $\boldsymbol{\Lambda}_2$ held fixed, we reach the right-hand side of (4·2) with

$$\boldsymbol{\Gamma}_p = (\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2), \quad \boldsymbol{\Lambda}_p = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 \end{pmatrix}, \quad \hat{\boldsymbol{\Lambda}}_p = \begin{pmatrix} \hat{\boldsymbol{\Lambda}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Lambda}}_2 \end{pmatrix}.$$

As before, $\hat{\boldsymbol{\Lambda}}_p$ consists of the diagonal elements of $\boldsymbol{\Gamma}_p^T \mathbf{X} \boldsymbol{\Gamma}_p$. Hence the minimum when $\boldsymbol{\Lambda}_2$ is unspecified is obtained by setting $\boldsymbol{\Lambda}_2 = \hat{\boldsymbol{\Lambda}}_2$, and

$$\min_H S(\boldsymbol{\Xi}) = \psi_{p_1+p_2}\{(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2)\} + \psi_{p_1}(\boldsymbol{\Lambda}_1 | \boldsymbol{\Gamma}_1).$$

If $H_0, H_1, \ldots, H_a$ is a nested sequence of hypotheses of the above type, where H_0 specifies $\boldsymbol{\Xi}$ completely and H_a leaves $\boldsymbol{\Xi}$ completely arbitrary, then we have an analysis of variance of the form

Source	*S.S.*	d.f.
About H_0 within H_1	$S(\boldsymbol{\Xi}_0) - S(\boldsymbol{\Xi}_1)$	$d_0 - d_1$
About H_1 within H_2	$S(\boldsymbol{\Xi}_1) - S(\boldsymbol{\Xi}_2)$	$d_1 - d_2$
...	...	...
About H_{a-1}	$S(\boldsymbol{\Xi}_{a-1})$	d_{a-1}
Total	$S(\boldsymbol{\Xi}_0)$	d_0

where $S(\boldsymbol{\Xi}_r) = \min_{H_r} S(\boldsymbol{\Xi})$ so that $S(\boldsymbol{\Xi}_a) = 0$, and if H_r specifies a_r l.v.'s and b_r of the corresponding l.r.'s, then

$$d_r = (k-1)\,a_r - \tfrac{1}{2}a_r(a_r - 1) + b_r.$$

4.3. *Wishart case.* Suppose $\mathbf{x}^T = (x_1, \ldots, x_k)$ is a vector variate having covariance matrix $\boldsymbol{\Xi}$. If $\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}$ are $n+1$ independent observations, then

$$\mathbf{X} = \frac{1}{n} \sum_{u=1}^{n+1} (\mathbf{x}_u - \bar{\mathbf{x}})(\mathbf{x}_u - \bar{\mathbf{x}})^T \quad \left(\bar{\mathbf{x}} = \frac{1}{n+1} \sum_{u=1}^{n+1} \mathbf{x}_u\right)$$

is the sample covariance matrix, and

$$\mathscr{E}(\mathbf{X}) = \boldsymbol{\Xi}.$$

The hypotheses described in § 3 have interpretations in terms of correlations between the basic variates. If these variates are assumed to be multinormal, the hypotheses refer to independence properties.

When $\boldsymbol{\Gamma}_p$ is given, let $\boldsymbol{\Gamma}_0$ be any $(k \times (k-p))$ matrix such that $(\boldsymbol{\Gamma}_p, \boldsymbol{\Gamma}_0)$ is orthogonal. Then the hypothesis $H_p(\boldsymbol{\Gamma}_p)$ specifies that the variates forming the elements of $\boldsymbol{\Gamma}_p^T \mathbf{x}$ are uncorrelated with each other and with the elements of $\boldsymbol{\Gamma}_0^T \mathbf{x}$; i.e. that they are principal components of the basic distribution. $H_p(\boldsymbol{\Gamma}_p; \boldsymbol{\Lambda}_p)$ specifies the variances of the elements of $\boldsymbol{\Gamma}_p^T \mathbf{x}$ also. In the normal case, the theory of testing these hypotheses is almost standard (see e.g. Anderson (1958) chapters 9 and 10); $n\mathbf{X}$ is now distributed in the Wishart form with n d.f., based on the covariance matrix $\boldsymbol{\Xi}$. Hence the likelihood ratio criterion for testing $H_p(\boldsymbol{\Gamma}_p)$ is given by

$$-2 \ln \lambda = n \ln \left\{ \prod_{i=1}^{p} (\boldsymbol{\Gamma}_p^T \mathbf{X} \boldsymbol{\Gamma}_p)_{ii} \, |\boldsymbol{\Gamma}_0^T \mathbf{X} \boldsymbol{\Gamma}_0| \, |\mathbf{X}|^{-1} \right\} \tag{4·7}$$

and that for testing $H_p(\boldsymbol{\Gamma}_p; \boldsymbol{\Lambda}_p)$ by

$$-2\ln\lambda = n\ln\{|\boldsymbol{\Lambda}_p|\,|\boldsymbol{\Gamma}_0^T\mathbf{X}\boldsymbol{\Gamma}_0|\,|\mathbf{X}|^{-1}\} + n\,\mathrm{tr}\,(\boldsymbol{\Gamma}_p^T\mathbf{X}\boldsymbol{\Gamma}_p\boldsymbol{\Lambda}_p^{-1}) - np. \tag{4·8}$$

The distributions of these criteria do not depend on any nuisance parameters; under the null hypothesis, they are asymptotically distributed as χ^2 with respectively d_p and $p+d_p$ d.f. Box's (1949) results show that a closer approximation for the first of these criteria is obtained by taking $\theta^{-1}(-2\ln\lambda)$ to be distributed as χ^2 with d_p d.f., where

$$\theta = 1 + \frac{1}{2n}\left(1+\frac{\alpha_3}{\alpha_2}\right), \quad \alpha_r = k^r - (k-p)^r - p.$$

However, his method does not apply to the second criterion, which has the same distribution as has

$$np\ln(ne^{-1}) - n\ln|\mathbf{A}_1| + \mathrm{tr}\,(\mathbf{A}_1) + \mathrm{tr}\,(\mathbf{A}_2),$$

where $\mathbf{A}_1(p\times p)$ and $\mathbf{A}_2\ (p\times p)$ are independently distributed in the Wishart form on respectively $n-k+p$ and $k-p$ d.f., based on the covariance matrix

$$\boldsymbol{\Sigma}_p = \boldsymbol{\Lambda}_p^{-\frac{1}{2}}\boldsymbol{\Gamma}_p^T\boldsymbol{\Xi}\boldsymbol{\Gamma}_p\boldsymbol{\Lambda}_p^{-\frac{1}{2}}$$

which reduces to $\mathbf{I}_p$ when the null hypothesis is true. The characteristic function of this criterion is easily found to be

$$\left(\frac{2e}{n}\right)^{-npit}\frac{|\boldsymbol{\Sigma}_p|^{-nit}}{|\mathbf{I}_p - 2it\boldsymbol{\Sigma}_p|^{\frac{1}{2}n-nit}}\prod_{j=1}^{p}\frac{\Gamma[\frac{1}{2}(n-k+p+1-j)-nit]}{\Gamma[\frac{1}{2}(n-k+p+1-j)]}$$

so that the mean is $\qquad p+d_p+n(\mathrm{tr}\,\boldsymbol{\Sigma}_p - p - \ln|\boldsymbol{\Sigma}_p|)+O(n^{-1}).$

4.4. *Rounding-off case.* If we follow through the likelihood ratio procedures assuming that the elements of $\mathbf{Z}$ are distributed uniformly on $(-\epsilon, +\epsilon)$, as is appropriate when they are rounding-off errors, we arrive in each case at test statistics which take only two values, one of which is zero; the non-zero value being attained if and only if there exists a matrix $\boldsymbol{\Xi}$ satisfying the constraints imposed by the hypothesis tested and such that each of the elements of $\mathbf{X}-\boldsymbol{\Xi}$ lies in $(-\epsilon, +\epsilon)$. The problem of determining whether such a matrix exists can be formulated as one of linear programming.

An alternative approach is to derive test statistics assuming that the elements of $\mathbf{Z}$ have the correct variances (as in (2·7)) but are normal instead of uniform, and then to find approximations to the null-hypothesis distributions of these statistics under the uniformity assumption. This leads to very complicated results except when $p=1$, in which case we obtain the statistics (for details of this derivation see T.R. 35)

$$\left.\begin{aligned}\psi_1^*(\boldsymbol{\gamma};\lambda) &= \psi_1^*(\boldsymbol{\gamma}) + \psi_1(\lambda|\boldsymbol{\gamma}),\\ \psi_1^*(\boldsymbol{\gamma}) &= \sigma^{-2}(\boldsymbol{\gamma}^T\mathbf{X}\mathbf{D}\mathbf{X}\boldsymbol{\gamma} - \hat{\lambda}^2\delta),\\ \psi_1^*(\lambda|\boldsymbol{\gamma}) &= \sigma^{-2}\delta(1+\delta)^{-1}(\lambda-\hat{\lambda})^2,\end{aligned}\right\} \tag{4·9}$$

where $\mathbf{D}(k\times k)$ is diagonal with $D_{ii} = (1-\gamma_i^2)^{-1}$, and

$$\delta = \boldsymbol{\gamma}^T\mathbf{D}\boldsymbol{\gamma}, \quad \hat{\lambda} = \delta^{-1}\boldsymbol{\gamma}^T\mathbf{D}\mathbf{X}\boldsymbol{\gamma}.$$

Under the normality assumption, these statistics have χ^2 distributions with respectively k, $k-1$, 1 d.f. Under the uniformity assumption, their expectations are unchanged, but their variances become complicated functions of the elements of $\boldsymbol{\gamma}$ (given explicitly in

T.R. 35, table 15). We give in Table 1 the values of these variances for the two types of vectors

$$\left.\begin{aligned} &\text{(i)} \quad \boldsymbol{\gamma}(\text{i}) = (0, \ldots, 0, 1, 0, \ldots, 0)^T \quad \text{(unit in any position)}, \\ &\text{(ii)} \quad \boldsymbol{\gamma}(\text{ii}) = k^{-\frac{1}{2}}(1, \pm 1, \ldots, \pm 1) \quad \text{(any allocation of signs)}. \end{aligned}\right\} \tag{4·10}$$

While this has not been proved rigorously, it seems likely that these vectors correspond to extreme values of the variances. It will be seen that for small k, with $\kappa_4 = -\frac{6}{5}$, these variances depend fairly strongly on the vector $\boldsymbol{\gamma}$. The simple assumption that the distributions are adequately represented by the normal theory χ^2 distributions is inadequate; a closer approximation is obtained by introducing a scale factor (depending on $\boldsymbol{\gamma}$) and modifying the d.f. However, the situation is still unsuited to the construction of confidence regions for $\boldsymbol{\gamma}$ or $(\boldsymbol{\gamma}; \lambda)$, since the statistics in (4·9) are complicated functions of $\boldsymbol{\gamma}$.

Table 1. *Means* (M) *and variances* (V) *of the statistics in* (4·9) *under the null hypothesis, assuming the elements of* $\mathbf{Z}$ *to have the rounding-off structure* (2·7), *for the two vectors in* (4·10)

Statistic	M	$V(\boldsymbol{\gamma}=\boldsymbol{\gamma}(\text{i}))$	$V(\boldsymbol{\gamma}=\boldsymbol{\gamma}(\text{ii}))$
$\psi_1^*(\boldsymbol{\gamma}; \lambda)$	k	$(2+\kappa_4)k$	$2k+\kappa_4\left(2-\frac{1}{k-1}+\frac{3}{(k-1)(2k-1)^2}\right)$
$\psi_1^*(\boldsymbol{\gamma})$	$k-1$	$(2+\kappa_4)(k-1)$	$2(k-1)+\kappa_4\left(2-\frac{5}{k}+\frac{2}{k(k-1)}\right)$
$\psi_1^*(\lambda\|\boldsymbol{\gamma})$	1	$2+\kappa_4$	$2+\kappa_4\left(\frac{8k-7}{k(2k-1)^2}\right)$

We therefore reject this approach, and suggest that the statistics appropriate to the invariant structure (given in (4·2), (4·3)) be used for the present case also. Assuming uniformity, the null-hypothesis moments of these statistics are easily found (see the next section); and the (approximate) confidence regions obtained using these statistics are much simpler to work with than are those obtained using (4·9).

5. NON-NORMAL MOMENTS OF THE NORMAL INVARIANT STATISTICS

We give some lower moments of the statistics given in (4·2), (4·3), (4·4), (4·5); first (case I), for the non-normal version of the invariant structure (2·5), and secondly (case R), for the rounding-off structure (2·7). The statistics for $1 < p < k$ are complex, and we have not calculated the second moments for these cases.

The moments of these statistics when $\mathbf{X}$ has a general Wishart distribution depend on the matrix $\boldsymbol{\Xi}$; when $\mathbf{X}$ has a unit Wishart distribution, we find

$$\mathscr{E}\{\psi_p(\boldsymbol{\Gamma}_p)\} = d_p, \quad \operatorname{var}\psi_p(\boldsymbol{\Gamma}_p) = 2d_p + \frac{2p}{n}(k^2 - \tfrac{1}{6}(p+1)(2p+1)).$$

For the rounding-off case with $\kappa_4 = -\frac{6}{5}$, and for each of the two 'extreme' vectors in (4·10), the means and variances given in Table 3 were evaluated numerically; equating these moments to those of $\theta\chi_f^2$ we arrive at the (approximate) 5 and 1 % points given in Table 5. It will be seen that the effect of $\boldsymbol{\gamma}$ on the distribution of $\psi_1(\boldsymbol{\gamma})$ and $\psi_1(\boldsymbol{\gamma}; \lambda)$ is considerable.

In Table 6 we give approximate 5 and 1 % points of distributions fitted to the moments of the statistics $\psi_k(\boldsymbol{\Gamma}_k)$ and $\psi_k(\boldsymbol{\Gamma}_k; \boldsymbol{\Lambda}_k)$, for the rounding-off case. The distribution of

$\psi_k(\mathbf{\Gamma}_k; \mathbf{\Lambda}_k)$ does not depend on $\mathbf{\Gamma}_k$; for $\psi_k(\mathbf{\Gamma}_k)$ we have taken, first, $\mathbf{\Gamma}_k = \mathbf{I}_k$, the unit $(k \times k)$ matrix, and secondly, we have averaged the moments in Table 4 by assuming all orientations of $\mathbf{\Gamma}_k$ to be equally likely (for details see T.R. 35). The effect of $\mathbf{\Gamma}_k$ is seen to be considerable.

Table 2. *Means of the statistics in* (4·2) *for the invariant* (*I*) *and rounding-off* (*R*) *structures* ((2·5) *and* (2·7), *respectively*); *p general*

Statistic	$M(I)$	$M(R)$
$\psi_p(\mathbf{\Gamma}_p)$	d_p	$d_p + \frac{1}{2}(\sum_{ab} \Gamma_{ab}^4 + \sum_{abc} \Gamma_{ab}^2 \Gamma_{ac}^2 - 2p)$
$\psi_p(\mathbf{\Lambda}_p \vert \mathbf{\Gamma}_p)$	p	$p - \frac{1}{2}\sum_{ab} \Gamma_{ab}^4$
$\psi_p(\mathbf{\Gamma}_p; \mathbf{\Lambda}_p)$	$p + d_p$	$p + d_p + \frac{1}{2}(\sum_{abc} \Gamma_{ab}^2 \Gamma_{ac}^2 - 2p)$

Table 3. *Means* (*M*) *and variances* (*V*) *of the statistics in* (4·3) *for the invariant* (*I*) *and rounding-off* (*R*) *structures*

Statistic	$M(I)$	$V(I)$	$M(R)$	$V(R)$
$\psi_1(\boldsymbol{\gamma})$	$k-1$	$2(k-1) + \frac{1}{2}\kappa_4(1 + (k-2)\Sigma_4 - 8\Sigma_6 + 8\Sigma_4^2)$	$k - 2 + \Sigma_4$	$2k - 6 + 6\Sigma_4 - 4\Sigma_6 + 2\Sigma_4^2 + \kappa_4(1 + (k-9)\Sigma_4 + 6\Sigma_6 - 7\Sigma_8 + 8\Sigma_4^2)$
$\psi_1(\lambda \vert \boldsymbol{\gamma})$	1	$2 + \kappa_4 \Sigma_4^2$	$1 - \frac{1}{2}\Sigma_4$	$2(1 - \frac{1}{2}\Sigma_4)^2 + \kappa_4(2\Sigma_4^2 - \frac{7}{4}\Sigma_8)$
$\psi_1(\boldsymbol{\gamma}; \lambda)$	k	$2k + \frac{1}{2}\kappa_4(1 + (k+2)\Sigma_4 - 4\Sigma_6 + 2\Sigma_4^2)$	$k - 1 + \frac{1}{2}\Sigma_4$	$2k - 4 + 4\Sigma_4 - 2\Sigma_6 + \frac{1}{2}\Sigma_4^2 + \kappa_4(1 + (k-5)\Sigma_4 + 3\Sigma_6 - \frac{7}{4}\Sigma_8 + 2\Sigma_4^2)$

(N.B. $\Sigma_r = \sum_a \gamma_a^r$).

Table 4. *Means* (*M*) *and variances* (*V*) *of the statistics in* (4·4) *and* (4·5) *for the invariant* (*I*) *and rounding-off* (*R*) *structures*; $p = k$

Statistic	$M(I)$	$V(I)$	$M(R)$	$V(R)$
$\psi_k(\mathbf{\Gamma}_k)$	d_k	$2d_k(1 + \frac{1}{2}\kappa_4) + \kappa_4(C - A)$	$\frac{1}{2}k(k-2) + \frac{1}{2}A$	$k(k - \frac{5}{2}) + A + \frac{1}{2}C + \kappa_4(\frac{1}{2}k^2 - \frac{9}{4}k + \frac{3}{2}A + 2C - \frac{7}{4}B)$
$\psi_k(\mathbf{\Lambda}_k \vert \mathbf{\Gamma}_k)$	k	$2k + \kappa_4 C$	$k - \frac{1}{2}A$	$2k - 2A + \frac{1}{2}C + \kappa_4(2C - \frac{7}{4}B)$
$\psi_k(\mathbf{\Gamma}_k; \mathbf{\Lambda}_k)$	$k + d_k$	$2(k + d_k)\left(1 + \frac{1}{4}\kappa_4 \frac{k+3}{k+1}\right)$	$\frac{1}{2}k^2$	$k(k - \frac{1}{2})(1 + \frac{1}{4}\kappa_4)$

(N.B. $A = \sum_a r_{aa}$, $B = \sum_a r_{aa}^2$, $C = \sum_{ab} r_{ab}^2$, $r_{ab} = \sum_c \Gamma_{ac}^2 \Gamma_{bc}^2$).

Table 5. *Approximate* 5 *and* 1 % *points of the statistics* $\psi_1(\boldsymbol{\gamma})$ and $\psi_1(\boldsymbol{\gamma}; \lambda)$ *in* (4·3) *in the rounding-off case, for the two vectors in* (4·10)

k	$\psi_1(\boldsymbol{\gamma}$ (i))		$\psi_1(\boldsymbol{\gamma}$ (ii))		$\psi_1(\boldsymbol{\gamma}$ (i); $\lambda)$		$\psi_1(\boldsymbol{\gamma}$ (ii); $\lambda)$	
2	2·77	4·13	1·71	2·81	3·43	4·73	3·20	4·61
3	4·43	6·03	3·70	5·50	5·04	6·62	5·02	6·96
4	5·92	7·72	5·48	7·76	6·50	8·29	6·79	9·18
5	7·32	9·28	7·18	9·83	7·90	9·85	8·48	11·25
6	8·68	10·78	8·80	11·78	9·25	11·35	10·09	13·20
7	10·00	12·23	10·37	13·63	10·56	12·80	11·66	15·04
8	11·29	13·64	11·89	15·41	11·85	14·21	13·17	16·81

Table 6. *Approximate 5 and 1 % points of the statistics* $\psi_k(\boldsymbol{\Gamma}_k; \boldsymbol{\Lambda}_k)$ *and* $\psi_k(\boldsymbol{\Gamma}_k)$ *in the rounding-off case*

k	$\psi_k(\boldsymbol{\Gamma}_k; \boldsymbol{\Lambda}_k)$		$\psi_k(\mathbf{I}_k)$		$\psi_k(\boldsymbol{\Gamma}_k)$, average	
2	4·82	6·78	2·77	4·13	2·43	3·88
3	8·81	11·45	5·92	7·72	5·34	7·29
4	13·75	17·01	10·00	12·23	9·13	11·56
5	19·66	23·53	15·06	17·73	13·85	16·74
6	26·57	31·02	21·11	24·22	19·52	22·86
7	34·46	39·51	28·16	31·70	26·16	29·94
8	43·35	48·98	36·21	40·17	33·78	38·01

6. CONFIDENCE REGIONS

6.1. *Normal case.* Having derived test criteria for the hypotheses of § 3, we can construct confidence regions for the p latent vectors comprising $\boldsymbol{\Gamma}_p$, or for the latent vectors and roots jointly. Thus, for example, assuming normality and the invariant covariance structure we have

$$P\{\psi_p(\boldsymbol{\Gamma}_p) < \chi^2_{d_p}(\alpha)\} = \alpha,$$

where $\chi^2_f(\alpha)$ denotes the $100\alpha\,\%$ point of the χ^2 distribution on f d.f. Hence the region in parameter space

$$\{\boldsymbol{\Gamma}_p : \psi_p(\boldsymbol{\Gamma}_p) < K_1\} \tag{6·1}$$

is a confidence region with confidence coefficient α if $K_1 = \chi^2_{d_p}(\alpha)$. If σ^2 is unknown, we replace it by its estimate s^2 and take $K_1 = d_p F_{d_p, f}(\alpha)$, where f is the d.f. of s^2.

6·2. *The case* $p = 1$. When $p = 1$, the (unit) vector $\boldsymbol{\gamma}$ can be represented as a point on the surface of the unit hypersphere (in k dimensions). If we impose the restriction (2·1), $\boldsymbol{\gamma}$ can vary only over half the surface of this hypersphere; to avoid non-essential complexities we shall relax this part of (2·1), and shall identify $\boldsymbol{\gamma}$ with a *pair* of points on the hypersphere, at opposite ends of a diameter. A confidence region for $\boldsymbol{\gamma}$ will then consist of a set of diametral point-pairs.

For almost all observed matrices $\mathbf{X}$, all the latent roots $l_1, \ldots, l_k$ of $\mathbf{X}$ are distinct, and the corresponding latent vectors $\mathbf{g}_1, \ldots, \mathbf{g}_k$ (columns of $\mathbf{G}$) are uniquely defined (either as points, subject to (2·1), or, as we shall assume, as diametral pairs of points on the unit hypersphere).

The $100\alpha\,\%$ confidence region for $\boldsymbol{\gamma}$ derived from $\psi_1(\boldsymbol{\gamma})$ is

$$R_N = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}^T\mathbf{X}^2\boldsymbol{\gamma} - (\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma})^2 < K_2\}, \tag{6·2}$$

where K_2 is $v\sigma^2\chi^2_{k-1}(\alpha)$ if σ^2 is known, or $(k-1)\,vs^2F_{k-1,f}(\alpha)$ if an estimate s^2 is used. For α not too near 1, R_N consists of k cap-pairs R_{Nj} $(j = 1, \ldots, k)$ on the unit hypersphere, where R_{Nj} contains $\mathbf{g}_j$. As α(or $v\sigma^2$) increases, the caps get larger, until they begin to coalesce; for α sufficiently large, R_N contains the whole surface of the hypersphere.

This last phenomenon is unattractive; in § 6.5 we discuss the possibility of constructing a system of confidence regions which do not suffer from this defect.

To find the shape of the cap-pair R_{Nj}, write

$$\mathbf{h} = \mathbf{G}^T\boldsymbol{\gamma}, \quad \boldsymbol{\gamma} = \mathbf{G}\mathbf{h}, \quad \mathbf{X}\mathbf{G} = \mathbf{G}\mathbf{L}, \quad \mathbf{L} = \operatorname{diag}(l_1, l_2, \ldots, l_k).$$

Then the region R_N becomes

$$R_N = \{\boldsymbol{\gamma} : \sum_{i(\neq j)} h_i^2(l_i - l_j)^2 - [\sum_{i(\neq j)} h_i^2(l_i - l_j)]^2 < K_2\} \quad (j = 1 \text{ or } 2 \text{ or } \ldots \text{ or } k). \tag{6·3}$$

If l_j is different from all the other l.r.'s (as it will be with probability 1), then for small values of K_2 the region R_{Nj} is approximately

$$R_{Nj} \cong R^*_{Nj} = \{\boldsymbol{\gamma}: \sum_{i(\neq j)} h_i^2(l_i - l_j)^2 < K_2\}$$
$$= \{\boldsymbol{\gamma}: \boldsymbol{\gamma}^T(\mathbf{X} - l_j\mathbf{I}_k)^2\boldsymbol{\gamma} < K_2\} \tag{6·4}$$

so that the boundary of R_{Nj} is given approximately by the intersection of the unit hypersphere with a certain hyperelliptic cylinder, having semi-axes

$$s_{ji} = K_2^{\frac{1}{2}}|l_i - l_j|^{-1} \quad (i \neq j).$$

The approximation involved in obtaining (6·4) will be adequate provided all these semi-axes are small compared with unity (the neglected terms in (6·4) are of order (s_{ji}^2) relative to those retained). A second approximation to the confidence coefficient associated with the region R^*_{Nj} in (6·4) is derived in § 7.

6·3. *Discussion.* The shape of the confidence region R_N is a novel feature of this problem, and merits some discussion. The statistic $\psi_1(\boldsymbol{\gamma})$ is a function of the matrix $\mathbf{X}$; let us write it temporarily as $\psi(\boldsymbol{\gamma}, \mathbf{X})$. We have shown that if $\boldsymbol{\gamma}$ is a l.v. of $\boldsymbol{\Xi}$, say $\boldsymbol{\gamma} = \boldsymbol{\gamma}_j$, then $\psi(\boldsymbol{\gamma}_j, \mathbf{X})$ is distributed as χ^2_{k-1}. This will be so for each of the l.v.'s; so for each j separately, the probability that the region

$$R_N = \{\boldsymbol{\gamma}: \psi(\boldsymbol{\gamma}, \mathbf{X}) < K_1\}$$

contains $\boldsymbol{\gamma}_j$ is the announced confidence coefficient α (if $K_1 = \chi^2_{k-1}(\alpha)$). Here, we can label the l.v.'s of $\boldsymbol{\Xi}$ in any way we choose, independent of $\mathbf{X}$; thus for example each of the following statements will be true with probability exactly α:

(i) R_N contains the l.v. of $\boldsymbol{\Xi}$ corresponding to the largest l.r.;

(ii) R_N contains the l.v. making the smallest angle with the vector $(1, 0, \ldots, 0)^T$.

The following questions now arise. Assuming that $v\sigma^2 K_1$ is sufficiently small, so that R_N consists of k cap-pairs, can we assert (with exactly specified confidence) that

(iii) of these k cap-pairs, a particular one (say that one containing the l.v. of $\mathbf{X}$ corresponding to the largest l.r. of $\mathbf{X}$) contains a l.v. of $\boldsymbol{\Xi}$; in particular, that

(iv) this cap-pair contains the l.v. of $\boldsymbol{\Xi}$ corresponding to the largest l.r. of $\boldsymbol{\Xi}$; or that

(v) each of the k cap-pairs contains a l.v. of $\boldsymbol{\Xi}$.

We can do none of these exactly, since it is never with probability 1 that R_N splits up into k cap-pairs; but, to a good approximation, if the latent roots of $\boldsymbol{\Xi}$ are well separated, (iii) and (iv) are approximate confidence statements with coefficient α (see § 7). With exact confidence we can assert a statement closely similar to (v), as follows:

(vi) R_N contains all the l.v.'s of $\boldsymbol{\Xi}$. The probability of this event is

$$P\{\psi(\boldsymbol{\gamma}_1, \mathbf{X}) < K_1, \ldots, \psi(\boldsymbol{\gamma}_k, \mathbf{X}) < K_1\} = P(k, K_1), \text{ say,}$$

which is not α^k, since the different ψ's are not independent; in fact they have the joint structure

$$\psi(\boldsymbol{\gamma}_j, \mathbf{X}) = \sum_{i \neq j} w_{ij}^2 \quad (j = 1, 2, \ldots, k),$$

where $\mathbf{W} = \boldsymbol{\Gamma}^T\mathbf{Z}\boldsymbol{\Gamma}$ is a symmetric matrix whose superdiagonal elements are independent unit normal variables. The explicit determination of the function $P(k, K_1)$ is difficult; however, heuristic reasoning suggests that the approximations

$$P(k, K_1) \cong \exp\{-k(1-\alpha)\} \cong 1 - k(1-\alpha)$$

(where $K_1 = \chi^2_{k-1}(\alpha)$) will be adequate over the usual range of values of P.

6.4. *Other cases.* The fact that we can thus derive from the statistic $\psi_1(\boldsymbol{\gamma})$ confidence statements involving all the l.v.'s of $\boldsymbol{\Xi}$ simultaneously, makes discussion of the regions obtained from the statistics $\psi_p(\boldsymbol{\Gamma}_p)$ for $1 < p \leqslant k$ somewhat superfluous. The regions in the cases $1 < p < k$ are in any event not easily visualized; in the case $p = k$ each of the three forms for $\psi_k(\boldsymbol{\Gamma}_k)$ given in (4·4) is useful in appreciating some aspects of the resulting region.

The joint confidence region for $(\boldsymbol{\gamma}; \lambda)$ derived from the statistic $\psi_1(\boldsymbol{\gamma}; \lambda)$ of (4·3) can be described as follows. The vector $\boldsymbol{\gamma}$ lies in a region of the same shape as R_N (6·2); for each such $\boldsymbol{\gamma}$, the corresponding λ lies in an interval centered on $\hat{\lambda} = \boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}$, with length

$$2d = 2\{(2v+c)\,\sigma^2(K_2 - \psi_1(\boldsymbol{\gamma}))\}^{\frac{1}{2}}. \tag{6·5}$$

The joint confidence coefficient will be α if K_2 is $\chi^2_k(\alpha)$. Again, we omit discussion of the cases $1 < p \leqslant k$.

6.5. *Further discussion.* We now consider the possibility of constructing a system of confidence or Bayes regions which will not suffer from the defect mentioned above with respect to the system R_N; we search for regions which for all $\mathbf{X}$ will leave part of the hypersphere uncovered, and which preferably will always consist of k separate cap-pairs. This discussion is somewhat academic, since the region R_N only fails to give k distinct cap-pairs when the latent roots of $\mathbf{X}$ are so close together (relative to $v\sigma^2$) that the corresponding l.v.'s are 'almost' indeterminate (for practical purposes). In this situation, the merging of several caps into a connected region can be regarded as a desirable property of the system R_N, since it indicates that several l.r.'s of $\boldsymbol{\Xi}$ may be close together (or even equal).

When σ^2 is known, the log-likelihood (4·1) is fully determined by $S(\boldsymbol{\Xi})$; but $\boldsymbol{\Xi}$ depends not only on $\boldsymbol{\gamma}$ but also on the other l.v.'s and on the l.r.'s, which are nuisance parameters. Thus we cannot apply an inverse probability argument to $\boldsymbol{\gamma}$ directly. Eliminating the nuisance parameters by minimizing $S(\boldsymbol{\Xi})$ with respect to them, we have

$$\min S(\boldsymbol{\Xi}) = \psi_1(\boldsymbol{\gamma})$$

so that the contours $\psi_1(\boldsymbol{\gamma}) = \text{const.}$ are contours of constant maximized likelihood. $\psi_1(\boldsymbol{\gamma})$ is a minimum (zero) when $\boldsymbol{\gamma}$ is any l.v. of $\mathbf{X}$; $\psi_1(\boldsymbol{\gamma})$ is a maximum ($(l_1 - l_k)^2/4v\sigma^2 = m$, say) when $\boldsymbol{\gamma} = (\boldsymbol{\xi}_1 \pm \boldsymbol{\xi}_k)/\sqrt{2}$.

In the confidence region argument used above, the bound K_1 in (6·1) depends on k and α only, and so is unrelated to the range of variation $(0, m)$ of $\psi_1(\boldsymbol{\gamma})$. A procedure taking account of the value of m (for example, one using the region (6·2) with $K_2 = K^*m$, K^* being a function of $v\sigma^2$, k, α) can be given exact confidence properties only if the distribution of m is known, and this depends on the l.r.'s of $\boldsymbol{\Xi}$. Similarly if instead of m some other function of the l.r.'s of $\mathbf{X}$ is used. The author has been unable to construct a system of exact confidence regions with the desired properties. (This is not to say that such a system cannot exist, or that a very good approximate system cannot be found.)

6.6. *Wishart case.* The confidence region R_W obtained by setting a bound K_3 on the statistic in (4·7) (with $p = 1$) is not readily described exactly; for K_3 not too large, it is composed of k cap-pairs on the unit hypersphere, of which a typical one is (to a first approximation)

$$R_{Wj} \cong \{\boldsymbol{\gamma}: \sum_{i(\neq j)} h_i^2(l_i - l_j)^2/l_i l_j < K_4\} \quad (\boldsymbol{\gamma} = \mathbf{G}\mathbf{h}).$$

For approximate confidence α, K_4 can be taken to be

$$K_4 = \{1 + (4n)^{-1}(3k+4)\}\,\chi^2_{k-1}(\alpha).$$

Thus the boundary of R_{Wj} is approximately elliptical, with semi-axes

$$s_{ji} = (K_4 l_i l_j)^{\frac{1}{2}} |l_i - l_j|^{-1} \quad (i \neq j).$$

Similarly, the statistic (4·8) (with $p = 1$) gives rise to a joint confidence region for $(\boldsymbol{\gamma}; \lambda)$. $\boldsymbol{\gamma}$ lies in a region of the same shape as R_W above; and for each such $\boldsymbol{\gamma}$, λ lies in an interval given by

$$n\left(\frac{\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}}{\lambda} - 1 + \ln\frac{\lambda}{\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}}\right) < K_5 - n\ln\left(\frac{\boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}\,|\boldsymbol{\Gamma}_0^T\mathbf{X}\boldsymbol{\Gamma}_0|}{|\mathbf{X}|}\right).$$

To the same degree of approximation as above, when $\boldsymbol{\gamma}$ is in the jth cap R_{Wj}, this interval is given by

$$n(\lambda - l_j)^2 < 2l_j^2(K_5 - n\sum_{i(\neq j)} h_i^2(l_i - l_j)^2/l_i l_j).$$

For approximate confidence α, K_5 can be taken to be $\chi^2_k(\alpha)$.

6.7. *Rounding-off case.* Here we shall obtain approximate confidence regions by setting bounds on the statistics in (4·3), which were derived originally for the normal invariant case; the values of the bounds will be obtained from the approximate distributions of these statistics in the rounding-off case, as characterized by the moments given in Table 3. Thus the confidence region will consist of k cap-pairs on the unit hypersphere; but in general the relative sizes of the caps will not be the same as in the normal case.

Let the l.v.'s of $\boldsymbol{\Xi}$ be $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k$. Then the means and variances of the statistics $\psi_1(\boldsymbol{\gamma}_j)$ $(j = 1, \ldots, k)$, namely (M_j, V_j) $(j = 1, \ldots, k)$, are given in Table 3 (with Σ_r here replaced by $\Sigma_{rj} = \sum_i \gamma_{ij}^r$ $(j = 1, \ldots, k)$). We take $\psi_1(\boldsymbol{\gamma}_j)$ to be distributed approximately as $\theta_j \chi^2_{f_j}$ where $M_j = \theta_j f_j$, $V_j = 2\theta_j^2 f_j$. In practice, we shall approximate these quantities by replacing $\boldsymbol{\gamma}_j$ by the corresponding l.v. $\mathfrak{g}_j$ of $\mathbf{X}$. Thus finally, the part of the desired approximate confidence region containing $\mathfrak{g}_j$ is

$$R_{Rj} = \{\boldsymbol{\gamma}: \sum_{i(\neq j)} h_i^2(l_i - l_j)^2 < K_j\} \quad (\boldsymbol{\gamma} = \mathbf{G}\mathbf{h}) \tag{6·6}$$

where for confidence α, we take K_j as $\sigma^2\theta_j\chi^2_{f_j}(\alpha)$. The discussion in section 6.3 applies here.

7. Further approximations

Lawley (1956) gave expansions in statistical differentials for the l.r.'s of a random symmetric matrix. The method extends trivially to give expressions for the l.v.'s also. We write as above

$$\mathbf{X} = \boldsymbol{\Xi} + \mathbf{Z}, \quad \boldsymbol{\Xi}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}, \quad \mathbf{X}\mathbf{G} = \mathbf{G}\mathbf{L},$$

$$\boldsymbol{\Gamma}^T\mathbf{X}\boldsymbol{\Gamma} = \boldsymbol{\Lambda} + \mathbf{W}, \quad \mathbf{W} = \boldsymbol{\Gamma}^T\mathbf{Z}\boldsymbol{\Gamma}, \quad \mathbf{H} = \mathbf{G}^T\boldsymbol{\Gamma},$$

where w_{ij} is $O(\epsilon)$, h_{ii} is $1 + O(\epsilon^2)$, h_{ij} is $O(\epsilon)$ for $i \neq j$. Then from the diagonal and off-diagonal elements of the equation

$$\mathbf{H}\boldsymbol{\Lambda} + \mathbf{H}\mathbf{W} = \mathbf{L}\mathbf{H}$$

we obtain

$$\left.\begin{aligned} h_{ii}(l_i - \lambda_i - w_{ii}) &= \sum_{a \neq i} h_{ia} w_{ai}, \\ h_{ij}(l_i - \lambda_j) &= h_{ii} w_{ij} + \sum_{a \neq i} h_{ia} w_{aj} \quad (i \neq j), \end{aligned}\right\} \tag{7·1}$$

whence

$$\left.\begin{aligned} l_i &= \lambda_i + w_{ii} + \sum_{a \neq i} \frac{w_{ai}^2}{\lambda_i - \lambda_a} + O(\epsilon^3), \\ (\lambda_i - \lambda_j) h_{ij} &= w_{ij} + \sum_{a \neq i} \frac{w_{ai} w_{aj}}{\lambda_i - \lambda_a} - \frac{w_{ii} w_{ij}}{\lambda_i - \lambda_j} + O(\epsilon^3) \quad (i \neq j), \\ h_{ii} &= 1 - \tfrac{1}{2} \sum_{a \neq i} \frac{w_{ai}^2}{(\lambda_i - \lambda_a)^2} + O(\epsilon^3) \end{aligned}\right\} \tag{7·2}$$

and by successive substitution in (7·1) the expansions can be taken as far as is desired.

These expansions will now be used to obtain a closer approximation to the confidence coefficient associated with the use of the region given (exactly) by the right-hand side of (6·4), in the normal invariant case. Then **W** has the same structure as has **Z** itself. Substituting from (7·2), we have

$$\sum_{i \neq j} h_{ij}^2 (l_i - l_j)^2 = \sum_{i \neq j} w_{ij}^2 + (l_j - \lambda_j - w_{jj})^2$$
$$= \sum_{i \neq j} w_{ij}^2 + \left(\sum_{i \neq j} \frac{w_{ij}^2}{\lambda_i - \lambda_j} \right)^2 + O(\epsilon^5).$$

Hence to this order,

$$P\{R_{Nj}^* \text{ contains } \boldsymbol{\gamma}_j\} = P\{x_j + v\sigma^2 y_j < K_1\} = P_j \text{ say},$$

where

$$K_1 = \chi_{k-1}^2(\alpha), \quad x_j = \sum_{i \neq j} u_{ij}^2, \quad y_j = (\sum_{i \neq j} c_{ij} u_{ij}^2)^2, \quad c_{ij}^{-1} = \lambda_i - \lambda_j, \quad u_{ij} = (v\sigma^2)^{-\frac{1}{2}} w_{ij},$$

so that the $k-1$ variables $\{u_{ij}\}$ $(i \neq j)$ are independent unit normal. Following the argument of Beale (1960) (pp. 65–66), and writing $p_f(K)$ for the probability density function of χ_f^2 evaluated at K, we have to the same order as before

$$P_j \cong \alpha - p_{k-1}(K_1) \mathscr{E}(v\sigma^2 y_j | x_j = K_1)$$
$$= \alpha - p_{k-1}(K_1) \frac{v\sigma^2 K_1^2}{k^2 - 1} C_j$$
$$= \alpha - v\sigma^2 C_j p_{k+3}(K_1),$$

where

$$C_j = (\sum_{i \neq j} c_{ij})^2 + 2 \sum_{i \neq j} c_{ij}^2. \tag{7·3}$$

Alternatively, to restore this probability (approximately) to the value α, we must replace the bound K_1 by

$$K_{1j}^* = K_1 \{1 + K_1 v\sigma^2 C_j (k^2 - 1)^{-1}\}. \tag{7·4}$$

In practice, we replace the c's by their estimates in terms of the l.r.'s of **X**. Thus finally, an approximate 100α % confidence region for $\boldsymbol{\gamma}$ consists of k cap-pairs on the unit hypersphere, the jth cap being

$$R_{Nj}^{**} = \{\boldsymbol{\gamma}: \boldsymbol{\gamma}^T (\mathbf{X} - l_j \mathbf{I}_k)^2 \boldsymbol{\gamma} < K_{1j}^* v\sigma^2\}, \tag{7·5}$$

where K_{1j}^* is given by (7·3) and (7·4), with

$$K_1 = \chi_{k-1}^2(\alpha), \quad c_{ij}^{-1} = l_i - l_j.$$

This cap is the intersection of the hypersphere with an elliptical hypercylinder, having semi-axes

$$s_{ji}^* = (K_{ij}^* v\sigma^2 c_{ij}^2)^{\frac{1}{2}} \quad (i \neq j)$$

The discussion in § 6.3 applies here.

If σ^2 is unknown but an independent estimate s^2 is available which is distributed as $\sigma^2 \chi_f^2 / f$, a similar argument shows that the jth cap of an approximate 100α % confidence region is given by (7·5), where now $K_{1j}^* v\sigma^2$ is replaced by $(k-1) v s^2 F_{1j}^*$, where

$$F_{1j}^* = F\left(1 + \frac{(k-1) v s^2 F}{k^2 - 1} \frac{f + k - 1}{f + (k-1) F} C_j\right) \tag{7·6}$$

and $F = F_{k-1,f}(\alpha)$.

8. Examples

The matrix

$$\mathbf{X} = \begin{pmatrix} 1{\cdot}6 & 1{\cdot}6 & 1{\cdot}2 \\ 1{\cdot}6 & 0{\cdot}3 & 0{\cdot}6 \\ 1{\cdot}2 & 0{\cdot}6 & 0{\cdot}7 \end{pmatrix}$$

was obtained by rounding off to one decimal a matrix the elements of which were approximately normal with variances of the order of $(0{\cdot}35)^2$ (actually the matrix (24·3) of T.R. 35). We shall use this matrix to give examples of the application of the methods of this paper both for the normal case and for the rounding-off case.

The latent roots and vectors of $\mathbf{X}$ are found to be given by

$$\mathbf{L} = \begin{pmatrix} 3{\cdot}3475 & 0 & 0 \\ 0 & 0{\cdot}0592 & 0 \\ 0 & 0 & -0{\cdot}8067 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 0{\cdot}7512 & 0{\cdot}2688 & 0{\cdot}6029 \\ 0{\cdot}4830 & 0{\cdot}3987 & -0{\cdot}7796 \\ 0{\cdot}4499 & -0{\cdot}8768 & -0{\cdot}1697 \end{pmatrix}.$$

8·1. *Normal case.* As explained in Mallows (1959), the matrix $\mathbf{X}$ (or rather the unrounded form of it) is the matrix of second-order regression coefficients (β_{ij}) obtained by fitting a quadratic response surface

$$f(\mathbf{y}) = \beta_0 + \sum_{i=1}^{3} \beta_i y_i + \sum_{i,j=1}^{3} \beta_{ij} y_i y_j \tag{8·1}$$

to data given by Box & Youle (1955). The experimental design used was central composite in $k = 3$ dimensions, but was not second-order rotatable; if it had been, then the invariant covariance structure (2·2) would have been appropriate. However, the departure from rotatability does not seem serious in this case, and we shall proceed on the assumption that it is negligible. Thus we shall assume (2·2) to hold, with

$$v = 0{\cdot}02539, \quad c = 0{\cdot}06197$$

(these quantities are characteristics of the design used). An estimate of σ^2, based on $f = 5$ d.f. is $s^2 = 4{\cdot}848$.

Box & Youle reduced the fitted surface to canonical form; this gives

$$\hat{f}(\mathbf{y}) = b_0 + \sum_i b_i y_i + \sum_i l_i h_i^2,$$

where $h_i = \mathbf{g}_i^T \mathbf{y}$, i.e. $\mathbf{h} = \mathbf{G}^T \mathbf{y}$ as before. The methods of this paper can be used to give confidence regions for the latent vectors of the matrix (β_{ij}), i.e. for the directions of the true canonical variables; and joint regions for the latent vectors and roots.

Using the statistic $\psi_1(\boldsymbol{\gamma})$ given in (4·3) with σ^2 replaced by its estimate s^2, we have for 95 % confidence the region

$$\{\boldsymbol{\gamma}: \boldsymbol{\gamma}^T \mathbf{X}^2 \boldsymbol{\gamma} - (\boldsymbol{\gamma}^T \mathbf{X} \boldsymbol{\gamma})^2 < (k-1) v s^2 F_1\}, \tag{8·2}$$

where $F_1 = F_{k-1,f}(\alpha) = 5{\cdot}79$ when $k = 3$, $f = 5$, $\alpha = 0{\cdot}95$.

An approximate region with exactly elliptical caps is given by (7·5) and (7·6); the cap R_{F1}^* centred on the first l.v. $\mathbf{g}_1$ of $\mathbf{X}$ (i.e. the first column of $\mathbf{G}$) is the intersection of the unit sphere with an elliptical cylinder with axis pointing along $\mathbf{g}_1$, and with semi-axes

$$s_{1i} = \{(k-1) v s^2 F_{11}^*\}^{\frac{1}{2}} |l_1 - l_i|^{-1} \quad (i = 2, 3), \tag{8·3}$$

$$\{s_{1i}\} = 0{\cdot}371, \quad 0{\cdot}294.$$

Here $F_{11}^* = F_1(1 + 0{\cdot}045) = 6{\cdot}05$.

These semi-axes point along the other two l.v.'s of $\mathbf{X}$. Similarly, the semi-axes of the other two caps R^*_{F2} and R^*_{F3} are found to be

$$\{s_{2i}\} = 0{\cdot}409, \quad 1{\cdot}55, \qquad \{s_{3i}\} = 0{\cdot}335, \quad 1{\cdot}605$$

the corresponding F^*'s being

$$F^*_{12} = 1{\cdot}27F_1 = 7{\cdot}35, \quad F^*_{13} = 1{\cdot}36\,F_1 = 7{\cdot}85.$$

8·2. *Discussion.* (i) Both of the semi-axes of R^*_{F1} are fairly small; the extremes of these axes correspond to angular deviations from $\mathbf{g}_1$ of 22° and 17°, respectively. The neglected term in (6·3) is at most 14 % of the term retained. Also the adjustment $F^*_{11} - F_1$ is small. We may therefore feel assured that this cap is a good approximation to a 95 % confidence region for the l.v. of Ξ associated with the largest l.r.

(ii) However R^*_{F2} and R^*_{F3} have each one semi-axis which exceeds unity; both of these lie in the same plane (orthogonal to $\mathbf{g}_1$). Obviously the approximation involved in deriving (7·6) is very suspect in these circumstances.

(iii) Investigating the exact 95 % confidence region (8·2), we find that the cap containing $\mathbf{g}_1$ is approximately elliptical; the semi-diameters in the directions of $\mathbf{g}_2$ and $\mathbf{g}_3$ (for comparison with the semi-axes $\{s_{1i}\}$ in (8·3)) are

$$0{\cdot}395, \quad 0{\cdot}301.$$

The rest of the region consists of a band round the sphere, of varying width.

(iv) In this problem it appeared likely *a priori* that two of the latent roots of Ξ are zero. Replacing l_2 and l_3 by zero in (8·3), we obtain as an approximate 95 % confidence region for the latent vector associated with the remaining latent root, a circular cap on the unit sphere, centred on $\mathbf{g}_1$, with radius

$$\{(k-1)\,vs^2F^*\}^{\frac{1}{2}}\,|l_1|^{-1} = 0{\cdot}366.$$

This value differs slightly from that given in T.R. 35, since there the unrounded version of $\mathbf{X}$ was used, and the correction $F^* - F$ was not applied.

(v) A joint confidence region for $(\boldsymbol{\gamma}; \lambda)$ can be obtained from the statistic $\psi_1(\boldsymbol{\gamma}; \lambda)$ in (4·3). With 95 % confidence, $\boldsymbol{\gamma}$ lies in a region of the same shape as (8·2), but with now

$$F_1 = F_{k,f}(\alpha) = 5{\cdot}41 \quad \text{with} \quad k = 3, \quad f = 5, \quad \alpha = 0{\cdot}95.$$

For each such $\boldsymbol{\gamma}$, the corresponding latent root λ lies in an interval centred on $\hat{\lambda} = \boldsymbol{\gamma}^T\mathbf{X}\boldsymbol{\gamma}$, with length given by (6·5) (with σ^2 replaced by its estimate, and $K_2 = kF(\alpha)$). Thus for $\boldsymbol{\gamma} = \mathbf{g}_1$, we have

$$\psi_1(\mathbf{g}_1) = 0, \quad \hat{\lambda} = l_1 = 3{\cdot}3475, \quad d = 2{\cdot}9785,$$

so that the interval for λ is (0·369, 6·326). The length of the interval for λ decreases to zero as $\boldsymbol{\gamma}$ approaches the boundary of its region.

8.3. *Rounding-off case.* For the purposes of an example, we shall now disregard the underlying normal variation in the elements of $\mathbf{X}$, and shall consider only the effect of the rounding-off errors. We assume these to be independent random variables, uniformly distributed in $(-0{\cdot}05, +0{\cdot}05)$. Thus the rounding-off structure (2·7) is appropriate, with

$$\sigma^2 = \tfrac{1}{3}(0{\cdot}05)^2 = 0{\cdot}000833, \quad \kappa_4 = -1{\cdot}2.$$

The construction of approximate confidence regions for the latent vectors is described in § 6.7; they consist of three cap-pairs; each of these being centred on a l.v. of $\mathbf{X}$ and having the same shape as in the normal case above, but with different relative sizes.

Taking first the cap containing the l.v. $\mathbf{g}_1$, we find

$$\Sigma_{41} = 0{\cdot}4138, \quad \Sigma_{61} = 0{\cdot}2007, \quad \Sigma_{81} = 0{\cdot}1060,$$

whence

$$\theta_1 = 0{\cdot}5673, \quad f_1 = 2{\cdot}4922.$$

Thus for 95 % confidence, we take in (6·6)

$$K_1\sigma^{-2} = \theta_1\chi^2_{f_1}(0{\cdot}95) = 3{\cdot}925$$

and the resulting cap is elliptical with semi-axes

$$\{s_{1i}\} = \{K_1^{\frac{1}{2}}|l_1 - l_i|^{-1}\} = 0{\cdot}0174, \quad 0{\cdot}0138.$$

Similarly for the 95 % caps containing $\mathbf{g}_2$ and $\mathbf{g}_3$, we find

$$\theta_2 = 0{\cdot}5766, \quad f_2 = 2{\cdot}8121, \quad \{s_{2i}\} = 0{\cdot}0182, \quad 0{\cdot}0693;$$
$$\theta_3 = 0{\cdot}5835, \quad f_3 = 2{\cdot}5746, \quad \{s_{3i}\} = 0{\cdot}0141, \quad 0{\cdot}0677.$$

Here we have not applied the adjustment in (7·4); strictly it is not applicable to the present non-normal case, but in any event the effect it is designed to counteract is here small compared with the approximation involved in using $\theta\chi^2_f$ to represent the distribution of $\psi_1(\boldsymbol{\gamma})$.

References

Anderson, T. W. (1958) *An introduction to multivariate statistical analysis.* New York: John Wiley & Sons.

Beale, E. M. L. (1960). Confidence regions in non-linear estimation. *J. R. Statist. Soc.* B, **22**, 41–76 (with discussion, pp. 76–88).

Box, G. E. P. (1949). A general distribution theory for a class of likelihood ratio criteria. *Biometrika*, **36**, 317–46.

Box, G. E. P. & Youle, P. V. (1955). The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. *Biometrics*, **11**, 287–323.

Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–36.

Mallows, C. L. (1959). Latent roots and vectors of random matrices. *Technical Report, No.* 35, Statistical Techniques Research Group, Princeton University. (Referred to as T.R. 35.)

Van der Vaart, H. R. (1958). Some results on the probability distribution of the latent roots of a symmetric matrix of continuously distributed elements, and some applications to the theory of response surface estimation. *Technical Report No.* 18, Institute of Statistics, University of North Carolina.

Biometrika (1961), **48**, 1 *and* 2, *p.* 151
Printed in Great Britain

Expected values of normal order statistics

BY H. LEON HARTER
Aeronautical Research Laboratory
Wright-Patterson Air Force Base

1. HISTORY

The problem of order statistics has received a great deal of attention from statisticians dating at least as far back as a paper by Karl Pearson (1902) giving a solution of a generalization of a problem proposed by Galton (1902). The generalized problem is that of finding the average difference between the pth and the $(p+1)$th individuals in a sample of size n when the sample is arranged in order of magnitude. The result is

$$\frac{n}{(n-p)!\,p!}\int_{-\infty}^{\infty}\alpha^{n-p}(1-\alpha)^p\,dx, \tag{1·1}$$

where $\alpha = \int_{-\infty}^{x}\phi(x)\,dx$ and $\phi(x)$ is the probability density function of the variable x. Pearson stated a theorem, which he attributed to W. F. Sheppard, that the average differences between successive individuals are the successive terms in the binomial expansion of

$$\int_{-\infty}^{\infty}\{\alpha+(1+\alpha)\}^n\,dx. \tag{1·2}$$

In a footnote, Pearson remarked, 'Clearly a knowledge of the average difference in character of two adjacent individuals involves also a knowledge of the average difference in character between any two individuals'. For a symmetric population, such knowledge also involves a knowledge of the expected values of all the order statistics, since for odd sample sizes $n = 2k+1$, where k is an integer, $E(x_{k+1}) = \mu$ (the population mean), while for even sample sizes $n = 2k$, $\frac{1}{2}[E(x_k)+E(x_{k+1})] = \mu$.

Irwin (1925) gave expressions for the mean difference between the pth and qth individuals in order of magnitude and for the moments of the frequency distribution of differences between consecutive individuals. Tippett (1925) published a seven-decimal-place table of the probability integral of the largest individual in samples of size n from a normal population for

$$n = 3, 5, 10 \text{ and } x = -2{\cdot}6\,(0{\cdot}2)\,5{\cdot}8;$$

$$n = 20, 30, 50 \text{ and } x = -0{\cdot}1\,(0{\cdot}1)\,6{\cdot}0; \quad n = 100\,(100)\,1000 \text{ and } x = 1{\cdot}0\,(0{\cdot}1)\,6{\cdot}5.$$

The same paper included a five-decimal-place table of the mean range of samples of size $n = 2\,(1)\,100$ from a normal population from which the expected values of the largest and smallest individuals could of course be derived. The expected values of normal order statistics other than the first and last were not computed until somewhat later.

Karl Pearson & Margaret V. Pearson (1931) obtained an expansion in Taylor series for $E(x_i)$, accurate to 5 or 6 decimal places for $|E(x_i)|$ not too large (say < 1). Fisher & Yates (1938, Table XX) published a two-decimal-place table of the expected values of all normal order statistics for samples of size $n = 2\,(1)\,50$. Their values are correct except for four errors of a unit in the last place, due to rounding. Hastings, Mosteller, Tukey & Winsor (1947)

published a five-decimal-place table of the means and standard deviations of all order statistics for samples of size $n = 2(1)10$ from a normal population, also from a uniform population and from a selected long-tailed population. Their values for the means of normal order statistics are correct except for $n = 10$, where there are errors of from 1 to 7 units in the last place.

Wilks (1948) published an expository paper summarizing work on order statistics up to that time and listing 90 references.

Godwin (1949*a*) published a table of the expected values of rank differences in normal samples, to 10 decimal places for $n = 2$; 9 decimal places for $n = 3, 4$; 8 decimal places for $n = 5$; 7 decimal places for $n = 6, 7$; 6 decimal places for $n = 8$; and 5 decimal places for $n = 9, 10$. Godwin (1949*b*) also published a seven-decimal-place table of the means and standard deviations of all normal order statistics for samples of size $n = 2(1)10$. His values for the means of the first-order statistics are accurate to 7 decimal places, and his other values are probally equally accurate, since they were computed by the same method. Cadwell (1953) published a table of moments (mean, variance, β_1 and β_2) and selected percentage points of the first quasi-range for samples of size $n = 10(1)30$. His values of the means are correct except for one error of a unit in the last place, due to rounding. E. S. Pearson & Hartley (1954, Table 28) published a table of expected values of normal order statistics, to 3 decimal places for $n = 2(1)20$ and to 2 decimal places for

$$n = 21(1)26(2)50;$$

values for $n = 2(1)10$ were compiled from Godwin's table, those for $n = 11(1)20$ were freshly computed by Jean H. Thompson, while those for $n > 20$ were taken from the table by Fisher and Yates. These values are correct except for three errors of a unit in the last place, due to rounding. Harter (1959) published a six-decimal-place table (accurate to within a unit in the last place) of the expected values of the range and of the first 8 quasi-ranges for samples of size $n = 2(1)100$ taken from a normal population. By dividing these values by two, the expectations of the absolute values of the nine largest and the nine smallest normal deviates can be obtained.

Federer (1951) used a somewhat different approach than did most of the aforementioned authors, who depended largely on numerical integration for the determination of tabular values. Federer made use of the recurrence formula

$$E(x_{m,i+1}) = \frac{1}{i}\{mE(x_{m-1,i}) - (m-i)E(x_{m,i})\}, \tag{1·3}$$

where $x_{m,i}$ is the ith largest deviate from a sample of size m. Starting from Tippett's table of expected values of the range, Federer computed three-decimal-place values of the three largest normal deviates for samples of size $n = 41(1)200$ and two-decimal-place values of the fourth largest normal deviate for $n = 41(1)200$ and of the fifth largest normal deviate for $n = 41(1)100$. Because of loss of accuracy with repeated application of the recurrence formula, some of Federer's values are in error by from 1 to 3 units in the last place, and it is evident that the form of the recurrence formula given by (1·3) is of little value in computation. The author is indebted to the Editor for pointing out that, if written in the form

$$E(x_{m-1,i}) = \frac{1}{m}\{iE(x_{m,i+1}) + (m-i)E(x_{m,i})\}, \tag{1·4}$$

the recurrence formula can be used for working *downwards* with no serious accumulation

of rounding errors. Similar recurrence formulae for the variance and covariance of order statistics have recently been obtained by Govindarajulu (1959).

2. Method of computation

The expected value of the kth largest observation in a sample of size n from a standard normal population (mean zero and variance one) is given by the equation

$$E(x_{k|n}) = \frac{n!}{(n-k)!\,(k-1)!}\int_{-\infty}^{\infty} x[\tfrac{1}{2}-\Phi(x)]^{k-1}[\tfrac{1}{2}+\Phi(x)]^{n-k}\,\phi(x)\,dx, \qquad (2{\cdot}1)$$

where $\phi(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$ and $\Phi(x) = \int_0^x \phi(x)\,dx$. The expected value of the kth smallest observation is given by the same expression preceded by a minus sign, so that for a given value of n it is necessary only to compute the expected values for $k = 1\,(1)\,[\frac{1}{2}n]$. This was done by numerical integration on the Univac Scientific (ERA 1103 A) computer, for $n = 2\,(1)\,100$ and for values of n, none of whose prime factors exceeds seven, up through $n = 400$. Values of $\log_{10} n!$ for $n = 1\,(1)\,400$ from a table by Pearson & Hartley (1954) and values of $\phi(x) = \phi(-x)$ for $x = 0\,(0{\cdot}05)\,7{\cdot}60$, $2\Phi(x) = -2\Phi(-x)$ for $x = 0\,(0{\cdot}05)\,5{\cdot}95$, and $1-2\Phi(x) = 1+2\Phi(-x)$ for $x = 6{\cdot}00\,(0{\cdot}05)\,7{\cdot}60$ from tables by the National Bureau of Standards (1953, Tables I and II) were read into the computer. For each pair of values of n and k, the product $I(n,k,x)$ of the multiplicative constant and the integrand was determined for $x = -7{\cdot}60\,(0{\cdot}05)\,7{\cdot}60$ by computing $e\log_e I(n,k,x)$, where

$$\log_e I(n,k,x) = \log_e n! - \log_e (n-k)! - \log_e (k-1)! + \log_e x + (k-1)\log_e[\tfrac{1}{2}-\Phi(x)] + (n-k)\log_e[\tfrac{1}{2}+\Phi(x)] + \log_e \phi(x). \qquad (2{\cdot}2)$$

Fixed-point binary arithmetic was used, and the numbers were scaled so as to retain as much accuracy as possible. Since $I(n,k,x)$ is zero (to the number of places carried in the computer) for all values of n and k when $|x| > 7{\cdot}60$, the resulting value of $E(x_{k|n})$, obtained by using either the trapezoidal rule or the seven-point Lagrangian integration formula, is found by summing $I(n,k,x)$ for $x = -7{\cdot}60\,(h)\,7{\cdot}60$ and multiplying by the interval, h. Results were computed and printed out (to seven decimal places) for $h = 0{\cdot}05$ and $h = 0{\cdot}10$, and agreement is sufficiently close to guarantee that the values of $E(x_{k|n})$ for $h = 0{\cdot}05$ are accurate to within a unit in the fifth decimal place. Accordingly, the values for $h = 0{\cdot}05$ were rounded to five decimal places, and the five-decimal-place values were punched on cards and printed on the IBM 407 tabulator. The results for $n = 2\,(1)\,100\,(25)\,250\,(50)\,400$ are shown in Table 1.

Acknowledgment, with thanks, is made to Eugene H. Guthrie, who programmed the problem for computation on the ERA 1103 A.

3. Blom's approximation

In 1954 Blom became interested in the problem of plotting points on normal probability paper and, after reading a paper by Chernoff & Lieberman (1954), in the related problem of estimating parameters by means of linear functions of order statistics, Blom (1958) proposed approximating the ith normal order statistic (ith smallest normal deviate) for a sample of size n by means of the relation

$$E(x_i) = \Phi^{-1}\left(\frac{i-\alpha}{n-2\alpha+1}\right), \qquad (3{\cdot}1)$$

where $\Phi(x) = \int_{-\infty}^{x} \phi(x)\,dx$, with $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$. Note that $\Phi(x)$ is defined differently here than in § 2. It should be mentioned that there has been an argument of long-standing between advocates of the approximations corresponding to $\alpha = 0$ and $\alpha = 0{\cdot}5$, neither of which is correct. Blom tabulated the value of α required to yield the correct value of $E(x_i)$ for $i = 1\,(1)\,[\frac{1}{2}n]$ when $n = 2\,(2)\,10\,(5)\,20$. The values of α increase as n increases, the lowest value being 0·330 for $n = 2$, $i = 1$. For a given n, α is least for $i = 1$, rises quickly to a peak

Table 2. *Values of* $\alpha_{i,n}$ *such that* $E(x_i) = \Phi^{-1}[(i-\alpha_{i,n})/(n-2\alpha_{i,n}+1)]$

i	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$	i	$n = 100$	$n = 200$	$n = 400$
1	0·377	0·384	0·391	0·396	0·401	30	0·394	0·404	0·414
2	·394	·403	·412	·419	·426	35	·393	·402	·412
3	·395	·405	·415	·423	·430	40	·392	·400	·410
4	·394	·405	·415	·424	·431	45	·391	·398	·408
5	·392	·403	·414	·423	·431	50	·391	·397	·407
6	0·391	0·402	0·412	0·422	0·430	55	—	0·396	0·405
7	·390	·400	·411	·421	·429	60	—	·395	·404
8	·389	·399	·410	·420	·429	65	—	·394	·403
9	·388	·398	·408	·418	·428	70	—	·394	·402
10	·388	·397	·407	·417	·427	75	—	·393	·401
11	0·387	0·396	0·406	0·416	0·426	80	—	0·393	0·400
12	·387	·395	·405	·415	·425	85	—	·392	·399
13	—	·394	·404	·414	·424	90	—	·392	·399
14	—	·393	·403	·414	·423	95	—	·391	·398
15	—	·393	·402	·413	·423	100	—	·391	·398
16	—	0·392	0·402	0·412	0·422	110	—	—	0·396
17	—	·392	·401	·411	·421	120	—	—	·396
18	—	·391	·400	·410	·420	130	—	—	·395
19	—	·391	·399	·410	·420	140	—	—	·394
20	—	·391	·399	·409	·419	150	—	—	·394
21	—	0·390	0·398	0·408	0·419	160	—	—	0·393
22	—	·390	·398	·408	·418	170	—	—	·393
23	—	·390	·397	·407	·417	180	—	—	·392
24	—	·390	·397	·407	·417	190	—	—	·392
25	—	·390	·396	·406	·416	200	—	—	·391

for a relatively small value of i, and then drops off slowly; as an example, for $n = 20$, $\alpha = 0{\cdot}374$ for $i = 1$, the peak value of α is 0·391 for $i = 3$, and α drops to 0·386 for $i = 8, 9, 10$. Blom conjectured that α always lies in the interval (0·33, 0·50). He suggested the use of $\alpha = \frac{3}{8}$ as a compromise value.

If one solves (3·1) for the value of α required to yield the correct value of $E(x_i)$ for given i and n, one obtains

$$\alpha_{i,n} = \frac{i-(n+1)\,\Phi[E(x_i)]}{1-2\Phi[E(x_i)]}. \tag{3·2}$$

Values of $\alpha_{i,n}$ for $i = 1\,(1)\,[\frac{1}{2}n]$ when $n = 25, 50, 100, 200, 400$ have been computed on the Burroughs E 101–3 computer, and the results, rounded to three decimal places, are shown in Table 2. For brevity, results have been given only for values of i which are multiples of 5 for i between 25 and 100 and multiples of 10 for i between 100 and 200. A glance at the values in Table 2 is sufficient to show that the compromise value of $\frac{3}{8}$ or 0·375 for α proposed by Blom

is too low except for small values of n. If, however, one wishes to minimize the maximum error in estimating $E(x_i)$ for $n \leqslant 400$, one is led to choose a value of α even small than $\frac{3}{8}$, since the estimate of $E(x_i)$ is much more sensitive to changes in α for small values of n (and i) than for large values. The maximum error in estimating $E(x_i)$ is minimized by choosing $\alpha = 0{\cdot}363$. This gives a maximum error of 0·018, which is hardly satisfactory. It is possible, however, to do a fairly good job of estimating $E(x_i)$ by choosing one or two compromise values of α for each n. One can choose a single compromise value, α_n, for each n, to be used for all values of i, and simultaneously insure that the error in $E(x_i)$ does not exceed four units in the third decimal place. If one uses $\alpha_{1,n}$ to estimate $E(x_1)$ and $\alpha_{2,n}$ to estimate $E(x_i)$ for $i \neq 1$, the error in $E(x_i)$ will not exceed one unit in the third decimal place. Values of α_n, $\alpha_{1,n}$ and $\alpha_{2,n}$ are given in Table 3 for $n = 2\,(2)\,10\,(5)\,25, 50, 100, 200, 400$ along with regression equations

Table 3. *Compromise values of* α

n	α_n	$\alpha_{1,n}$	$\alpha_{2,n}$
2	0·330	0·330	—
4	·349	·347	0·359
6	·359	·355	·368
8	·364	·360	·374
10	0·368	0·364	0·378
15	·374	·370	·385
20	·378	·374	·390
25	·381	·377	·394
50	0·389	0·384	0·403
100	·396	·391	·412
200	·402	·396	·419
400	·407	·401	·426

To estimate α for intermediate values of n, use the following equations:

$$\alpha_n = 0{\cdot}314195 + 0{\cdot}063336X - 0{\cdot}010895X^2,$$
$$\alpha_{1,n} = 0{\cdot}315065 + 0{\cdot}057974X - 0{\cdot}009776X^2,$$
$$\alpha_{2,n} = 0{\cdot}327511 + 0{\cdot}058212X - 0{\cdot}007909X^2,$$
where $X = \log_{10} n$.

to be used for intermediate values of n. Values of α found by substituting tabular values of n in these regression equations do not differ from the corresponding tabular values of α by more than two units in the third decimal place, and this error in α does not increase the error in $E(x_i)$ by more than one unit in the third decimal place. There is reason to believe that results for intermediate values of n will be equally good, but use of these equations for $n > 400$ is emphatically discouraged. Thus, if one wishes to interpolate for intermediate values of n, the maximum errors are two units in the third decimal place for the approximation based on a single compromise value of α and five units in the third decimal place for the approximation based on two compromise values of α. These errors compare with a maximum error of between one and two units in the third decimal place for linear interpolation between successive value of n for a given $i(k)$ in Table 1. Comparison of the maximum errors might lead to the conclusion that interpolation in Table 1 is always more accurate than interpolation using Blom's approximation This would be erroneous, since the maximum error for the former occurs for large values of i (near $\frac{1}{2}n$), while the maximum error for the latter occurs for small values of i. Interpolation using Blom's approximation for large values of i, especially when the desired value of n lies about midway between widely separated successive tabular values of n (for example, when $n = 232$), and interpolation in Table 1 otherwise will limit the error to no more than a unit in the third decimal place. If more accurate values are required, they should be computed in the same way that Table 1 was computed, as should values for $n > 400$, or else they should be computed by

working downwards from the next higher tabular value of n, using the recurrence formula (1·4). Table 4 summarizes the above results, giving maximum errors in determining $E(x_i)$ by various methods.

Table 4. *Maximum errors in determining $E(x_i)$ by various methods*

Method	Values of n in Table 3	Intermediate values of n
Blom's approximation:		
$\alpha = 0{\cdot}363$ for all values of n	0·018	0·018
One value of α for each n	·004	·005
Two values of α for each n	·001	·002
Interpolation in Table 1	—	·002
Recurrence formula (1·4)	·00001	·00001
Numerical integration ($h = 0{\cdot}05$)	$<$·00001	$<$·00001

4. APPLICATIONS

Pearson & Hartley (1954, p. 56) have given two examples of applications of tables of expected values of normal order statistics. The first of these is concerned with estimating the weight of the five heaviest of 30 lambs at age $2\frac{1}{2}$ months, given the mean and standard deviation of the population, which is assumed to be normal. The second deals with the use of order statistics in estimating the population standard deviation. Pearson & Hartley and also Fisher & Yates (1953, p. 76) mention the use of expected values of normal order statistics in the analysis of variance of ranked data. The potential use of expected values of normal order statistics for transformation to standard normal scores preliminary to the analysis of variance was the principal motivation for the present study. In cases where only the rank of the observations is known, there is no reasonable alternative to transformation to standard normal scores, but the usefulness of this method is not restricted to such cases. When the data are known to have come from a population which does not satisfy the assumptions underlying the analysis of variance, of which normality is one, or when the data themselves give a strong indication to that effect, the experimenter seeks a transformation which will minimize or eliminate departures from the assumptions. One transformation which should be considered is the transformation to standard normal scores, and a preliminary investigation has shown that this transformation has some very desirable properties; in some cases it reduces both non-additivity and non-homogeneity of variance to lower levels than does any transformation of the form $(x+c)^p$. It has, of course, the obvious disadvantage of not being reversible.

REFERENCES

BLOM, GUNNAR (1958). *Statistical Estimates and Transformed Beta-Variables.* New York: John Wiley and Sons, Inc.

CADWELL, J. H. (1953). The distribution of quasi-ranges in samples from a normal population. *Ann. Math. Statist.* **24**, 603–13.

CHERNOFF, HERMAN & LIEBERMAN, GERALD J. (1954). Use of normal probability paper. *J. Amer. Statist. Ass.* **49**, 778–85.

FEDERER, WALTER T. (1951). *Evaluation of Variance Components from a Group of Experiments with Multiple Classifications.* Iowa Agricultural Experiment Station Research Bulletin No. 380.

FISHER, RONALD A. & YATES, FRANK (1953). *Statistical Tables for Biological, Agricultural and Medical Research*, 4th edition (1st edition, 1938). Edinburgh: Oliver and Boyd.

GALTON, FRANCIS (1902). The most suitable proportion between the values of first and second prizes. *Biometrika*, **1**, 385–90.

GODWIN, H. J. (1949*a*). On the estimation of dispersion by linear systematic statistics. *Biometrika*, **36**, 92–100.

GODWIN, H. J. (1949*b*). Some low moments of order statistics. *Ann. Math. Statist.* **20**, 279–85.

GOVINDARAJULU, Z. (1959). On moments of order statistics from normal populations (abstract). *Ann. Math. Statist.* **30**, 617.

HARTER, H. LEON (1959). The use of sample quasi-ranges in estimating population standard deviation. *Ann. Math. Statist.* **30**, 980–99.

HASTINGS, CECIL, JR., MOSTELLER, FREDERICK, TUKEY, JOHN W. & WINSOR, CHARLES, P. (1947). Low moments for small samples: a comparative study of order statistics. *Ann. Math. Statist.* **18**, 413–26.

IRWIN, J. O. (1925). The further theory of Francis Galton's individual difference problem. *Biometrika*, **17**, 100–28.

NATIONAL BUREAU OF STANDARDS (1953). *Tables of Normal Probability Functions.* Applied Mathematics Series No. 23.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press for the Biometrika Trustees.

PEARSON, KARL (1902). Note on Francis Galton's problem. *Biometrika*, **1**, 390–9.

PEARSON, KARL & PEARSON, MARGARET, V. (1931). On the mean character and variance of a ranked individual, and on the mean and variance of the intervals between ranked individuals. *Biometrika*, **23**, 364–97.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364–87.

WILKS, S. S. (1948). Order statistics. *Bull. Amer. Math. Soc.* **54**, 6–50.

ADDENDUM

An account of methods of computing expected values of normal order statistics would be incomplete without mention of the series expansions worked out by David & Johnson (1954) and by Plackett (1958). Saw (1960) has made a comparison of the David–Johnson series and the Plackett series. Neither series seems particularly well adapted to the computation of tables of the sort included in this paper, though either would be quite useful in obtaining very accurate expected values for isolated cases. The author wishes to thank Dr F. N. David for drawing his attention to these papers.

REFERENCES

DAVID, F. N. & JOHNSON, N. L. (1954). Statistical treatment of censored data. *Biometrika*, **41**, 228–40.

PLACKETT, R. L. (1958). Linear estimation from censored data. *Ann. Math. Statist.* **29**, 131–42.

SAW, J. G. (1960). A note on the error after the number of terms of the David–Johnson series for the expected values of normal order statistics. *Biometrika*, **47**, 79–86.

Table 1. *Expected values of normal order statistics* (see overleaf)

$$E(x_{k|n}) = \frac{n!}{(n-k)!\,(k-1)!}\int_{-\infty}^{\infty} x[\tfrac{1}{2}-\Phi(x)]^{k-1}[\tfrac{1}{2}+\Phi(x)]^{n-k}\,\phi(x)\,dx,$$

where $$\phi(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2} \quad \text{and} \quad \Phi(x) = \int_0^x \phi(x)\,dx.$$

[Tabular values are the expected values of the kth largest normal deviate for a sample of size n from $N(0, 1)$; or when preceded by a minus sign, they are the expected values of the kth smallest normal deviate.]

k \ n			2	3	4	5	6	7	8	9
1			0·56419	0·84628	1·02938	1·16296	1·26721	1·35218	1·42360	1·48501
2			—	·00000	0·29701	0·49502	0·64176	0·75737	0·85222	0·93230
3			—	—	—	·00000	·20155	·35271	·47282	·57197
4			—	—	—	—	—	·00000	·15251	·27453
5			—	—	—	—	—	—	—	·00000

k \ n	10	11	12	13	14	15	16	17	18	19
1	1·53875	1·58644	1·62923	1·66799	1·70338	1·73591	1·76599	1·79394	1·82003	1·84448
2	1·00136	1·06192	1·11573	1·16408	1·20790	1·24794	1·28474	1·31878	1·35041	1·37994
3	0·65606	0·72884	0·79284	0·84983	0·90113	0·94769	0·99027	1·02946	1·06573	1·09945
4	·37576	·46198	·53684	·60285	·66176	·71488	·76317	0·80738	0·84812	0·88586
5	·12267	·22489	·31225	·38833	·45557	·51570	·57001	·61946	·66479	·70661
6	—	0·00000	0·10259	0·19052	0·26730	0·33530	0·39622	0·45133	0·50158	0·54771
7	—	—	—	·00000	0·08816	·16530	·23375	·29519	·35084	·40164
8	—	—	—	—	—	·00000	·07729	·14599	·20774	·26374
9	—	—	—	—	—	—	—	·00000	·06880	·13072
10	—	—	—	—	—	—	—	—	—	·00000

k \ n	20	21	22	23	24	25	26	27	28	29
1	1·86748	1·88917	1·90969	1·92916	1·94767	1·96531	1·98216	1·99827	2·01371	2·02852
2	1·40760	1·43362	1·45816	1·48137	1·50338	1·52430	1·54423	1·56326	1·58145	1·59888
3	1·13095	1·16047	1·18824	1·21445	1·23924	1·26275	1·28511	1·30641	1·32674	1·34619
4	0·92098	0·95380	0·98459	1·01356	1·04091	1·06679	1·09135	1·11471	1·13697	1·15822
5	·74538	·78150	·81527	0·84697	0·87682	0·90501	0·93171	0·95705	0·98115	1·00414
6	0·59030	0·62982	0·66667	0·70115	0·73354	0·76405	0·79289	0·82021	0·84615	0·87084
7	·44833	·49148	·53157	·56896	·60399	·63690	·66794	·69727	·72508	·75150
8	·31493	·36203	·40559	·44609	·48391	·51935	·55267	·58411	·61385	·64205
9	·18696	·23841	·28579	·32965	·37047	·40860	·44436	·47801	·50977	·53982
10	·06200	·11836	·16997	·21755	·26163	·30268	·34105	·37706	·41096	·44298
11	—	0·00000	0·05642	0·10813	0·15583	0·20006	0·24128	0·27983	0·31603	0·35013
12	—	—	—	·00000	·05176	·09953	·14387	·18520	·22389	·26023
13	—	—	—	—	—	·00000	·04781	·09220	·13361	·17240
14	—	—	—	—	—	—	—	·00000	·04442	·08588
15	—	—	—	—	—	—	—	—	—	·00000

k \ n	30	31	32	33	34	35	36	37	38	39
1	2·04276	2·05646	2·06967	2·08241	2·09471	2·10661	2·11812	2·12928	2·14009	2·15059
2	1·61560	1·63166	1·64712	1·66200	1·67636	1·69023	1·70362	1·71659	1·72914	1·74131
3	1·36481	1·38268	1·39985	1·41637	1·43228	1·44762	1·46244	1·47676	1·49061	1·50402
4	1·17855	1·19803	1·21672	1·23468	1·25196	1·26860	1·28466	1·30016	1·31514	1·32964
5	1·02609	1·04709	1·06721	1·08652	1·10509	1·12295	1·14016	1·15677	1·17280	1·18830
6	0·89439	0·91688	0·93841	0·95905	0·97886	0·99790	1·01624	1·03390	1·05095	1·06741
7	·77666	·80066	·82359	·84555	·86660	·88681	0·90625	0·92496	0·94300	0·96041
8	·66885	·69438	·71875	·74204	·76435	·78574	·80629	·82605	·84508	·86343
9	·56834	·59545	·62129	·64596	·66954	·69214	·71382	·73465	·75468	·77398
10	·47329	·50206	·52943	·55552	·58043	·60427	·62710	·64902	·67009	·69035
11	0·38235	0·41287	0·44185	0·46942	0·49572	0·52084	0·54488	0·56793	0·59005	0·61131
12	·29449	·32686	·35755	·38669	·41444	·44091	·46620	·49042	·51363	·53592
13	·20885	·24322	·27573	·30654	·33582	·36371	·39032	·41576	·44012	·46348
14	·12473	·16126	·19572	·22832	·25924	·28863	·31663	·34336	·36892	·39340
15	·04148	·08037	·11695	·15147	·18415	·21515	·24463	·27272	·29954	·32520
16	—	0·00000	0·03890	0·07552	0·11009	0·14282	0·17388	0·20342	0·23159	0·25849
17	—	—	—	·00000	·03663	·07123	·10399	·13509	·16469	·19292
18	—	—	—	—	—	·00000	·03461	·06739	·09853	·12817
19	—	—	—	—	—	—	—	·00000	·03280	·06395
20	—	—	—	—	—	—	—	—	—	·00000

k \ n	40	41	42	43	44	45	46	47	48	49
1	2·16078	2·17068	2·18032	2·18969	2·19882	2·20772	2·21639	2·22486	2·23312	2·24119
2	1·75312	1·76458	1·77571	1·78654	1·79707	1·80733	1·81732	1·82706	1·83655	1·84582
3	1·51702	1·52964	1·54188	1·55377	1·56533	1·57658	1·58754	1·59820	1·60860	1·61874
4	1·34368	1·35728	1·37048	1·38329	1·39574	1·40784	1·41962	1·43108	1·44224	1·45312
5	1·20330	1·21782	1·23190	1·24556	1·25881	1·27170	1·28422	1·29641	1·30827	1·31983
6	1·08332	1·09872	1·11364	1·12810	1·14213	1·15576	1·16899	1·18186	1·19439	1·20658
7	0·97722	0·99348	1·00922	1·02446	1·03924	1·05358	1·06751	1·08104	1·09420	1·10701
8	·88114	·89825	0·91480	0·93082	0·94634	0·96139	0·97599	0·99018	1·00396	1·01737
9	·79259	·81056	·82792	·84472	·86097	·87673	·89201	·90684	0·92125	0·93525
10	·70988	·72871	·74690	·76448	·78148	·79795	·81391	·82939	·84442	·85902
11	0·63177	0·65149	0·67052	0·68889	0·70666	0·72385	0·74049	0·75663	0·77228	0·78748
12	·55736	·57799	·59788	·61707	·63561	·65353	·67088	·68768	·70397	·71978
13	·48591	·50749	·52827	·54830	·56763	·58631	·60438	·62186	·63881	·65523
14	·41688	·43944	·46114	·48204	·50220	·52166	·54046	·55865	·57625	·59331
15	·34978	·37337	·39604	·41784	·43885	·45912	·47868	·49759	·51588	·53360
16	0·28423	0·30890	0·33257	0·35533	0·37723	0·39833	0·41868	0·43834	0·45734	0·47573
17	·21988	·24569	·27043	·29418	·31701	·33898	·36016	·38060	·40034	·41942
18	·15644	·18345	·20931	·23411	·25792	·28081	·30285	·32410	·34460	·36441
19	·09362	·12192	·14897	·17488	·19972	·22358	·24652	·26862	·28992	·31049
20	·03117	·06085	·08917	·11625	·14219	·16707	·19097	·21396	·23610	·25746
21	—	0·00000	0·02969	0·05803	0·08513	0·11109	0·13600	0·15993	0·18296	0·20514
22	—	—	—	·00000	·02835	·05546	·08144	·10637	·13033	·15338
23	—	—	—	—	—	·00000	·02712	·05311	·07805	·10203
24	—	—	—	—	—	—	—	·00000	·02599	·05095
25	—	—	—	—	—	—	—	—	—	·00090

k \ n	50	51	52	53	54	55	56	57	58	59
1	2·24907	2·25678	2·26432	2·27169	2·27891	2·28598	2·29291	2·29970	2·30635	2·31288
2	1·85487	1·86371	1·87235	1·88080	1·88906	1·89715	1·90506	1·91282	1·92041	1·92786
3	1·62863	1·63829	1·64773	1·65695	1·66596	1·67478	1·68340	1·69185	1·70012	1·70822
4	1·46374	1·47409	1·48420	1·49407	1·50372	1·51315	1·52237	1·53140	1·54024	1·54889
5	1·33109	1·34207	1·35279	1·36326	1·37348	1·38346	1·39323	1·40278	1·41212	1·42127
6	1·21846	1·23003	1·24132	1·25234	1·26310	1·27361	1·28387	1·29391	1·30373	1·31334
7	1·11948	1·13162	1·14347	1·15502	1·16629	1·17729	1·18804	1·19855	1·20882	1·21886
8	1·03042	1·04312	1·05550	1·06757	1·07934	1·09083	1·10205	1·11300	1·12371	1·13419
9	0·94887	0·96213	0·97504	0·98762	0·99988	1·01185	1·02352	1·03493	1·04607	1·05695
10	·87321	·88701	·90045	·91354	·92629	0·93873	0·95086	0·96271	0·97427	0·98557
11	0·80225	0·81661	0·83058	0·84417	0·85742	0·87033	0·88292	0·89520	0·90719	0·91890
12	·73513	·75004	·76455	·77866	·79240	·80578	·81883	·83155	·84397	·85609
13	·67117	·68660	·70170	·71633	·73057	·74444	·75794	·77111	·78396	·79649
14	·60986	·62592	·64152	·65668	·67143	·68578	·69976	·71337	·72665	·73960
15	·55077	·56742	·58358	·59928	·61455	·62940	·64385	·65793	·67164	·68502
16	0·49354	0·51080	0·52755	0·54380	0·55960	0·57495	0·58989	0·60444	0·61860	0·63241
17	·43789	·45578	·47312	·48995	·50629	·52217	·53761	·55263	·56725	·58150
18	·38357	·40211	·42007	·43749	·45439	·47080	·48675	·50226	·51736	·53205
19	·33036	·34957	·36818	·38621	·40369	·42065	·43713	·45314	·46872	·48388
20	·27807	·29799	·31726	·33592	·35400	·37154	·38856	·40510	·42117	·43681
21	0·22653	0·24719	0·26716	0·28648	0·30518	0·32331	0·34090	0·35797	0·37456	0·39068
22	·17559	·19702	·21772	·23772	·25708	·27583	·29400	·31163	·32875	·34538
23	·12511	·14735	·16880	·18953	·20957	·22896	·24774	·26595	·28362	·30078
24	·07494	·09803	·12029	·14177	·16252	·18259	·20201	·22082	·23906	·25677
25	·02496	·04896	·07206	·09434	·11584	·13661	·15669	·17614	·19498	·21325
26	—	0·00000	0·02400	0·04712	0·06940	0·09091	0·11170	0·13180	0·15127	0·17013
27	—	—	—	·00000	·02312	·04541	·06693	·08773	·10785	·12733
28	—	—	—	—	—	·00000	·02229	·04382	·06463	·08476
29	—	—	—	—	—	—	—	·00000	·02153	·04234
30	—	—	—	—	—	—	—	—	—	·00000

k \ n	60	61	62	63	64	65	66	67	68	69
1	2·31928	2·32556	2·33173	2·33778	2·34373	2·34958	2·35532	2·36097	2·36652	2·37199
2	1·93516	1·94232	1·94934	1·95624	1·96301	1·96965	1·97618	1·98260	1·98891	1·99510
3	1·71616	1·72394	1·73158	1·73906	1·74641	1·75363	1·76071	1·76767	1·77451	1·78122
4	1·55736	1·56567	1·57381	1·58180	1·58963	1·59732	1·60487	1·61228	1·61955	1·62670
5	1·43023	1·43900	1·44760	1·45603	1·46430	1·47241	1·48036	1·48817	1·49584	1·50338
6	1·32274	1·33195	1·34097	1·34982	1·35848	1·36698	1·37532	1·38351	1·39154	1·39942
7	1·22869	1·23832	1·24774	1·25698	1·26603	1·27490	1·28360	1·29213	1·30051	1·30873
8	1·14443	1·15445	1·16427	1·17388	1·18329	1·19252	1·20157	1·21044	1·21915	1·22769
9	1·06760	1·07802	1·08821	1·09819	1·10797	1·11754	1·12693	1·13613	1·14516	1·15401
10	0·99662	1·00742	1·01799	1·02833	1·03846	1·04838	1·05810	1·06762	1·07696	1·08612
11	0·93034	0·94153	0·95247	0·96317	0·97365	0·98391	0·99395	1·00380	1·01345	1·02291
12	·86793	·87950	·89081	·90187	·91270	·92329	·93367	0·94383	0·95379	0·96355
13	·80873	·82068	·83237	·84379	·85496	·86590	·87660	·88708	·89735	·90741
14	·75224	·76459	·77665	·78843	·79996	·81123	·82226	·83306	·84364	·85400
15	·69807	·71081	·72324	·73540	·74727	·75889	·77025	·78138	·79226	·80293
16	0·64587	0·65901	0·67183	0·68436	0·69659	0·70856	0·72025	0·73170	0·74290	0·75387
17	·59538	·60893	·62214	·63504	·64764	·65996	·67200	·68377	·69529	·70657
18	·54637	·56033	·57395	·58723	·60020	·61288	·62526	·63737	·64921	·66080
19	·49864	·51303	·52705	·54073	·55408	·56712	·57985	·59230	·60447	·61638
20	·45202	·46685	·48129	·49537	·50911	·52252	·53561	·54841	·56091	·57314
21	0·40637	0·42164	0·43652	0·45101	0·46515	0·47894	0·49240	0·50555	0·51839	0·53095
22	·36155	·37729	·39260	·40752	·42207	·43625	·45009	·46360	·47680	·48969
23	·31745	·33366	·34944	·36480	·37976	·39435	·40857	·42245	·43601	·44925
24	·27396	·29066	·30691	·32272	·33812	·35312	·36775	·38201	·39594	·40953
25	·23098	·24820	·26494	·28122	·29706	·31249	·32753	·34219	·35649	·37045
26	0·18842	0·20618	0·22343	0·24019	0·25650	0·27237	0·28784	0·30290	0·31759	0·33192
27	·14621	·16452	·18230	·19957	·21636	·23269	·24859	·26408	·27917	·29389
28	·10425	·12315	·14148	·15927	·17656	·19337	·20973	·22565	·24116	·25627
29	·06248	·08198	·10089	·11923	·13704	·15435	·17118	·18755	·20349	·21902
30	·02081	·04096	·06047	·07938	·09774	·11556	·13288	·14972	·16611	·18207
31	—	0·00000	0·02014	0·03966	0·05858	0·07694	0·09478	0·11211	0·12896	0·14536
32	—	—	—	·00000	·01952	·03844	·05681	·07465	·09199	·10885
33	—	—	—	—	—	·00000	·01893	·03730	·05514	·07249
34	—	—	—	—	—	—	—	·00000	·01837	·03622
35	—	—	—	—	—	—	—	—	—	·00000

k \ n	70	71	72	73	74	75	76	77	78	79
1	2·37736	2·38265	2·38785	2·39298	2·39802	2·40299	2·40789	2·41271	2·41747	2·42215
2	2·00120	2·00720	2·01310	2·01890	2·02462	2·03024	2·03578	2·04124	2·04662	2·05191
3	1·78783	1·79432	1·80071	1·80699	1·81317	1·81926	1·82525	1·83115	1·83696	1·84268
4	1·63373	1·64063	1·64742	1·65410	1·66067	1·66714	1·67350	1·67976	1·68592	1·69200
5	1·51078	1·51805	1·52520	1·53223	1·53914	1·54594	1·55263	1·55921	1·56569	1·57207
6	1·40717	1·41478	1·42226	1·42961	1·43684	1·44395	1·45094	1·45782	1·46459	1·47125
7	1·31680	1·32473	1·33252	1·34017	1·34770	1·35510	1·36237	1·36953	1·37657	1·38350
8	1·23608	1·24431	1·25240	1·26034	1·26815	1·27583	1·28338	1·29080	1·29810	1·30529
9	1·16270	1·17123	1·17961	1·18784	1·19592	1·20387	1·21168	1·21936	1·22691	1·23434
10	1·09511	1·10393	1·11259	1·12110	1·12945	1·13766	1·14572	1·15365	1·16145	1·16912
11	1·03220	1·04130	1·05024	1·05902	1·06764	1·07610	1·08442	1·09260	1·10063	1·10854
12	0·97313	0·98252	0·99173	1·00078	1·00966	1·01838	1·02695	1·03537	1·04364	1·05178
13	·91728	·92695	·93644	0·94576	0·95490	0·96387	0·97269	0·98135	0·98986	0·99822
14	·86416	·87412	·88388	·89346	·90286	·91209	·92115	·93005	·93880	·94739
15	·81338	·82362	·83366	·84351	·85317	·86265	·87196	·88110	·89008	·89890
16	0·76462	0·77514	0·78546	0·79558	0·80550	0·81524	0·82480	0·83418	0·84339	0·85244
17	·71761	·72843	·73903	·74942	·75960	·76960	·77940	·78903	·79848	·80776
18	·67214	·68325	·69413	·70480	·71526	·72551	·73557	·74544	·75512	·76463
19	·62803	·63943	·65060	·66155	·67227	·68279	·69310	·70322	·71314	·72289
20	·58510	·59681	·60827	·61950	·63050	·64128	·65185	·66222	·67239	·68237
21	0·54323	0·55525	0·56701	0·57852	0·58980	0·60085	0·61168	0·62230	0·63272	0·64294
22	·50230	·51463	·52669	·53850	·55006	·56138	·57248	·58336	·59403	·60449
23	·46219	·47484	·48721	·49932	·51117	·52277	·53414	·54528	·55621	·56692
24	·42281	·43579	·44848	·46089	·47304	·48493	·49657	·50798	·51917	·53013
25	·38404	·39739	·41041	·42313	·43558	·44777	·45970	·47138	·48283	·49404

k \ n	70	71	72	73	74	75	76	77	78	79
26	0·34591	0·35958	0·37292	0·38597	0·39873	0·41122	0·42343	0·43540	0·44711	0·45859
27	·30825	·32227	·33596	·34934	·36242	·37521	·38772	·39997	·41196	·42371
28	·27102	·28540	·29945	·31317	·32657	·33968	·35250	·36504	·37731	·38934
29	·23416	·24893	·26333	·27740	·29114	·30457	·31770	·33055	·34311	·35542
30	·19762	·21277	·22756	·24199	·25608	·26984	·28329	·29645	·30931	·32190
31	0·16134	0·17690	0·19208	0·20688	0·22133	0·23543	0·24922	0·26269	0·27586	0·28875
32	·12527	·14125	·15683	·17202	·18684	·20130	·21543	·22923	·24272	·25591
33	·08936	·10579	·12178	·13737	·15257	·16740	·18188	·19602	·20983	·22334
34	·05357	·07045	·08688	·10289	·11848	·13370	·14854	·16303	·17718	·19101
35	·01785	·03520	·05209	·06852	·08453	·10014	·11536	·13021	·14471	·15888
36	—	0·00000	0·01736	0·03424	0·05068	0·06670	0·08231	0·09754	0·11240	0·12691
37	—	—	—	·00000	·01689	·03333	·04935	·06497	·08020	·09507
38	—	—	—	—	—	·00000	·01644	·03247	·04809	·06333
39	—	—	—	—	—	—	—	·00000	·01602	·03165
40	—	—	—	—	—	—	—	—	—	·00000

k \ n	80	81	82	83	84	85	86	87	88	89
1	2·42677	2·43133	2·43582	2·44026	2·44463	2·44894	2·45320	2·45741	2·46156	2·46565
2	2·05714	2·06228	2·06735	2·07236	2·07729	2·08216	2·08696	2·09170	2·09637	2·10099
3	1·84832	1·85387	1·85935	1·86475	1·87007	1·87532	1·88049	1·88560	1·89064	1·89561
4	1·69798	1·70387	1·70968	1·71540	1·72104	1·72660	1·73209	1·73750	1·74283	1·74810
5	1·57836	1·58455	1·59065	1·59665	1·60258	1·60841	1·61417	1·61984	1·62544	1·63096
6	1·47781	1·48428	1·49064	1·49691	1·50309	1·50918	1·51518	1·52110	1·52693	1·53269
7	1·39032	1·39704	1·40366	1·41017	1·41659	1·42292	1·42915	1·43529	1·44135	1·44732
8	1·31236	1·31932	1·32617	1·33292	1·33957	1·34611	1·35257	1·35893	1·36520	1·37138
9	1·24165	1·24884	1·25593	1·26290	1·26977	1·27653	1·28320	1·28976	1·29624	1·30262
10	1·17666	1·18409	1·19139	1·19859	1·20567	1·21264	1·21951	1·22628	1·23295	1·23952
11	1·11631	1·12396	1·13148	1·13889	1·14618	1·15336	1·16043	1·16740	1·17426	1·18102
12	1·05978	1·06764	1·07539	1·08300	1·09050	1·09788	1·10515	1·11231	1·11936	1·12631
13	1·00644	1·01453	1·02249	1·03031	1·03802	1·04560	1·05306	1·06041	1·06765	1·07478
14	0·95584	0·96414	0·97231	0·98034	0·98825	0·99603	1·00369	1·01122	1·01865	1·02596
15	·90757	·91609	·92447	·93271	·94082	·94880	0·95665	0·96437	0·97198	0·97948
16	0·86134	0·87007	0·87867	0·88711	0·89542	0·90360	0·91164	0·91956	0·92735	0·93502
17	·81687	·82583	·83464	·84329	·85180	·86017	·86841	·87651	·88449	·89234
18	·77398	·78315	·79217	·80103	·80975	·81832	·82675	·83504	·84320	·85123
19	·73246	·74186	·75109	·76016	·76908	·77785	·78647	·79496	·80330	·81152
20	·69217	·70179	·71124	·72053	·72965	·73862	·74744	·75611	·76465	·77304
21	0·65297	0·66282	0·67249	0·68199	0·69133	0·70050	0·70952	0·71838	0·72710	0·73568
22	·61476	·62484	·63473	·64445	·65399	·66337	·67259	·68165	·69056	·69932
23	·57742	·58773	·59785	·60779	·61755	·62714	·63656	·64581	·65492	·66387
24	·54088	·55143	·56178	·57193	·58191	·59171	·60133	·61079	·62009	·62923
25	·50504	·51583	·52641	·53680	·54700	·55701	·56684	·57650	·58600	·59533
26	0·46985	0·48088	0·49170	0·50232	0·51274	0·52297	0·53301	0·54288	0·55258	0·56210
27	·43522	·44651	·45757	·46842	·47907	·48952	·49979	·50986	·51976	·52949
28	·40111	·41265	·42397	·43506	·44594	·45662	·46710	·47739	·48750	·49743
29	·36747	·37927	·39084	·40218	·41330	·42421	·43491	·44542	·45574	·46587
30	·33423	·34630	·35813	·36972	·38108	·39223	·40316	·41389	·42443	·43477
31	0·30136	0·31371	0·32580	0·33765	0·34926	0·36065	0·37182	0·38278	0·39353	0·40409
32	·26881	·28144	·29381	·30592	·31779	·32943	·34084	·35203	·36300	·37378
33	·23655	·24947	·26212	·27450	·28664	·29852	·31018	·32161	·33281	·34381
34	·20453	·21775	·23069	·24335	·25576	·26790	·27981	·29148	·30292	·31415
35	·17272	·18625	·19949	·21244	·22512	·23753	·24970	·26162	·27330	·28476
36	0·14108	0·15493	0·16848	0·18172	0·19469	0·20738	0·21981	0·23199	0·24392	0·25562
37	·10959	·12377	·13763	·15118	·16444	·17741	·19012	·20256	·21475	·22669
38	·07820	·09272	·10691	·12078	·13434	·14761	·16059	·17330	·18576	·19796
39	·04689	·06177	·07629	·09049	·10436	·11793	·13121	·14420	·15692	·16938
40	·01562	·03087	·04575	·06028	·07448	·08836	·10193	·11521	·12821	·14094
41	—	0·00000	0·01524	0·03013	0·04466	0·05886	0·07275	0·08633	0·09961	0·11262
42	—	—	—	·00000	0·01488	·02942	·04362	·05751	·07110	·08439
43	—	—	—	—	—	·00000	·01454	·02874	·04263	·05622
44	—	—	—	—	—	—	—	·00000	·01421	·02810
45	—	—	—	—	—	—	—	—	—	·00000

Table 1 (*cont.*)

k \ n	90	91	92	93	94	95	96	97	98	99
1	2·46970	2·47370	2·47764	2·48154	2·48540	2·48920	2·49297	2·49669	2·50036	2·50400
2	2·10554	2·11004	2·11448	2·11887	2·12321	2·12749	2·13172	2·13590	2·14003	2·14411
3	1·90052	1·90536	1·91015	1·91487	1·91953	1·92414	1·92869	1·93318	1·93763	1·94201
4	1·75329	1·75842	1·76348	1·76848	1·77341	1·77828	1·78309	1·78784	1·79254	1·79718
5	1·63641	1·64178	1·64709	1·65232	1·65749	1·66259	1·66763	1·67261	1·67752	1·68238
6	1·53836	1·54396	1·54949	1·55494	1·56033	1·56564	1·57089	1·57607	1·58118	1·58624
7	1·45321	1·45903	1·46476	1·47042	1·47600	1·48151	1·48695	1·49232	1·49762	1·50286
8	1·37747	1·38348	1·38941	1·39526	1·40103	1·40673	1·41235	1·41790	1·42338	1·42879
9	1·30891	1·31511	1·32123	1·32726	1·33321	1·33909	1·34489	1·35061	1·35626	1·36183
10	1·24600	1·25239	1·25869	1·26491	1·27104	1·27708	1·28305	1·28894	1·29475	1·30049
11	1·18769	1·19426	1·20073	1·20712	1·21342	1·21964	1·22577	1·23182	1·23779	1·24368
12	1·13316	1·13990	1·14656	1·15311	1·15958	1·16596	1·17226	1·17847	1·18459	1·19064
13	1·08181	1·08873	1·09555	1·10228	1·10891	1·11546	1·12191	1·12827	1·13455	1·14075
14	1·03316	1·04026	1·04726	1·05415	1·06095	1·06765	1·07426	1·08078	1·08721	1·09356
15	0·98686	0·99413	1·00129	1·00835	1·01531	1·02217	1·02894	1·03561	1·04219	1·04868
16	0·94258	0·95002	0·95735	0·96458	0·97170	0·97872	0·98564	0·99246	0·99919	1·00583
17	·90007	·90769	·91519	·92258	·92986	·93704	·94411	·95109	·95797	0·96475
18	·85914	·86693	·87460	·88215	·88959	·89693	·90416	·91129	·91831	·92524
19	·81960	·82756	·83540	·84312	·85072	·85822	·86560	·87288	·88006	·88713
20	·78131	·78944	·79745	·80533	·81310	·82075	·82829	·83572	·84305	·85027
21	0·74412	0·75243	0·76061	0·76866	0·77659	0·78441	0·79210	0·79968	0·80716	0·81452
22	·70795	·71643	·72478	·73300	·74110	·74907	·75692	·76466	·77228	·77980
23	·67267	·68134	·68986	·69825	·70651	·71464	·72266	·73055	·73832	·74598
24	·63822	·64706	·65576	·66432	·67275	·68105	·68922	·69727	·70519	·71301
25	·60451	·61353	·62241	·63115	·63974	·64821	·65654	·66474	·67282	·68079
26	0·57147	0·58068	0·58974	0·59865	0·60742	0·61605	0·62454	0·63291	0·64115	0·64926
27	·53905	·54845	·55769	·56678	·57572	·58452	·59318	·60170	·61010	·61837
28	·50718	·51677	·52620	·53547	·54459	·55356	·56239	·57108	·57963	·58805
29	·47582	·48561	·49522	·50468	·51398	·52312	·53212	·54097	·54969	·55827
30	·44493	·45491	·46472	·47436	·48384	·49316	·50233	·51136	·52024	·52898
31	0·41445	0·42463	0·43464	0·44447	0·45414	0·46364	0·47299	0·48218	0·49123	0·50013
32	·38436	·39474	·40495	·41498	·42483	·43452	·44404	·45341	·46263	·47170
33	·35461	·36520	·37561	·38584	·39588	·40576	·41547	·42501	·43440	·44364
34	·32517	·33598	·34660	·35702	·36727	·37733	·38722	·39695	·40652	·41593
35	·29601	·30704	·31787	·32850	·33895	·34921	·35929	·36920	·37895	·38853
36	0·26710	0·27835	0·28940	0·30025	0·31090	0·32136	0·33163	0·34173	0·35166	0·36142
37	·23841	·24990	·26117	·27223	·28309	·29375	·30423	·31452	·32464	·33458
38	·20991	·22164	·23314	·24443	·25550	·26637	·27705	·28754	·29785	·30797
39	·18159	·19356	·20530	·21681	·22810	·23919	·25008	·26077	·27127	·28159
40	·15341	·16563	·17761	·18936	·20088	·21219	·22328	·23418	·24488	·25539
41	0·12536	0·13783	0·15006	0·16205	0·17380	0·18533	0·19665	0·20776	0·21866	0·22937
42	·09740	·11014	·12262	·13486	·14685	·15861	·17015	·18148	·19259	·20351
43	·06952	·08253	·09528	·10777	·12001	·13201	·14378	·15533	·16666	·17778
44	·04169	·05499	·06801	·08076	·09325	·10550	·11750	·12928	·14083	·15217
45	·01389	·02748	·04078	·05381	·06656	·07906	·09131	·10332	·11510	·12666
46	—	0·00000	0·01359	0·02689	0·03992	0·05267	0·06518	0·07743	0·08944	0·10123
47	—	—	—	·00000	·01330	·02633	·03909	·05159	·06385	·07586
48	—	—	—	—	—	·00000	·01303	·02579	·03829	·05055
49	—	—	—	—	—	—	—	·00000	·01276	·02527
50	—	—	—	—	—	—	—	—	—	·00000

Table 1 (*cont.*)

k \ n	100	125	150	175	200	225	250	300	350	400
1	2·50759	2·58634	2·64925	2·70148	2·74604	2·78485	2·81918	2·87777	2·92651	2·96818
2	2·14814	2·23630	2·30638	2·36434	2·41365	2·45649	2·49431	2·55867	2·61207	2·65761
3	1·94635	2·04090	2·11578	2·17755	2·22999	2·27547	2·31555	2·38365	2·44004	2·48806
4	1·80176	1·90146	1·98019	2·04500	2·09991	2·14746	2·18932	2·26033	2·31904	2·36897
5	1·68718	1·79137	1·87341	1·94081	1·99783	2·04713	2·09050	2·16397	2·22462	2·27615
6	1·59123	1·69947	1·78448	1·85419	1·91308	1·96395	2·00864	2·08427	2·14663	2·19955
7	1·50803	1·62002	1·70777	1·77959	1·84019	1·89247	1·93837	2·01595	2·07985	2·13402
8	1·43414	1·54966	1·63997	1·71376	1·77594	1·82953	1·87654	1·95592	2·02122	2·07654
9	1·36734	1·48623	1·57896	1·65462	1·71828	1·77310	1·82115	1·90220	1·96882	2·02521
10	1·30615	1·42828	1·52333	1·60075	1·66583	1·72182	1·77084	1·85348	1·92133	1·97871
11	1·24950	1·37477	1·47206	1·55118	1·61760	1·67470	1·72466	1·80879	1·87781	1·93614
12	1·19661	1·32493	1·42438	1·50514	1·57287	1·63103	1·68189	1·76746	1·83758	1·89681
13	1·14687	1·27819	1·37975	1·46210	1·53109	1·59027	1·64199	1·72894	1·80013	1·86021
14	1·09982	1·23409	1·33771	1·42161	1·49182	1·55200	1·60455	1·69283	1·76504	1·82595
15	1·05509	1·19226	1·29791	1·38333	1·45472	1·51588	1·56923	1·65880	1·73201	1·79371
16	1·01238	1·15243	1·26007	1·34697	1·41953	1·48163	1·53577	1·62659	1·70076	1·76324
17	0·97145	1·11435	1·22396	1·31232	1·38602	1·44904	1·50395	1·59599	1·67109	1·73432
18	·93208	1·07783	1·18937	1·27917	1·35399	1·41792	1·47359	1·56681	1·64283	1·70678
19	·89411	1·04268	1·15616	1·24738	1·32330	1·38812	1·44452	1·53891	1·61582	1·68048
20	·85739	1·00879	1·12417	1·21680	1·29381	1·35950	1·41663	1·51216	1·58994	1·65530
21	0·82179	0·97601	1·09330	1·18731	1·26540	1·33195	1·38980	1·48645	1·56508	1·63112
22	·78720	·94426	1·06344	1·15883	1·23798	1·30539	1·36393	1·46169	1·54116	1·60786
23	·75353	·91342	1·03449	1·13126	1·21146	1·27971	1·33895	1·43780	1·51809	1·58544
24	·72070	·88344	1·00639	1·10452	1·18577	1·25485	1·31478	1·41470	1·49580	1·56379
25	·68863	·85423	0·97907	1·07855	1·16084	1·23074	1·29135	1·39233	1·47423	1·54285
26	0·65725	0·82573	0·95245	1·05329	1·13661	1·20733	1·26861	1·37063	1·45332	1·52257
27	·62651	·79789	·92650	1·02868	1·11303	1·18457	1·24651	1·34957	1·43303	1·50289
28	·59635	·77065	·90115	1·00469	1·09005	1·16240	1·22500	1·32908	1·41332	1·48378
29	·56672	·74398	·87638	0·98125	1·06763	1·14079	1·20405	1·30914	1·39414	1·46520
30	·53758	·71782	·85212	·95835	1·04574	1·11970	1·18361	1·28971	1·37546	1·44711
31	0·50890	0·69215	0·82836	0·93594	1·02434	1·09909	1·16365	1·27076	1·35725	1·42948
32	·48062	·66692	·80506	·91399	1·00340	1·07895	1·14415	1·25225	1·33947	1·41228
33	·45273	·64212	·78219	·89247	0·98290	1·05923	1·12507	1·23415	1·32211	1·39550
34	·42518	·61770	·75973	·87135	·96279	1·03992	1·10640	1·21646	1·30515	1·37910
35	·39796	·59365	·73764	·85062	·94307	1·02098	1·08810	1·19914	1·28854	1·36306
36	0·37102	0·56993	0·71590	0·83025	0·92371	1·00241	1·07016	1·18217	1·27229	1·34736
37	·34436	·54653	·69450	·81022	·90469	0·98418	1·05256	1·16553	1·25637	1·33199
38	·31793	·52343	·67341	·79051	·88599	·96626	1·03528	1·14921	1·24076	1·31693
39	·29173	·50061	·65261	·77110	·86760	·94866	1·01830	1·13320	1·22544	1·30216
40	·26572	·47804	·63210	·75197	·84950	·93134	1·00161	1·11746	1·21041	1·28767
41	0·23990	0·45571	0·61185	0·73312	0·83167	0·91429	0·98520	1·10200	1·19565	1·27344
42	·21423	·43361	·59184	·71453	·81410	·89751	·96905	1·08680	1·18114	1·25947
43	·18870	·41172	·57208	·69618	·79678	·88098	·95314	1·07185	1·16688	1·24574
44	·16330	·39002	·55253	·67806	·77969	·86469	·93748	1·05713	1·15285	1·23225
45	·13800	·36851	·53319	·66016	·76283	·84862	·92204	1·04264	1·13904	1·21897
46	0·11279	0·34717	0·51405	0·64247	0·74619	0·83277	0·90682	1·02836	1·12545	1·20590
47	·08765	·32598	·49509	·62498	·72975	·81712	·89180	1·01429	1·11207	1·19304
48	·06257	·30494	·47632	·60768	·71350	·80168	·87699	1·00042	1·09888	1·18037
49	·03753	·28403	·45770	·59056	·69744	·78642	·86236	0·98674	1·08587	1·16789
50	·01251	·26325	·43925	·57361	·68156	·77134	·84792	·97324	1·07305	1·15559

Table 1 (*cont.*)

k \ n	125	150	175	200	225	250	300	350	400
51	0·24258	0·42094	0·55682	0·66585	0·75644	0·83365	0·95991	1·06041	1·14346
52	·22201	·40278	·54019	·65030	·74170	·81955	·94676	1·04793	1·13149
53	·20154	·38475	·52371	·63490	·72712	·80561	·93376	1·03561	1·11969
54	·18115	·36684	·50737	·61966	·71270	·79183	·92093	1·02345	1·10804
55	·16084	·34904	·49116	·60456	·69842	·77819	·90824	1·01144	1·09654
56	0·14059	0·33136	0·47508	0·58959	0·68428	0·76470	0·89570	0·99957	1·08518
57	·12040	·31378	·45913	·57476	·67028	·75135	·88329	·98784	1·07396
58	·10026	·29630	·44329	·56005	·65641	·73812	·87102	·97624	1·06287
59	·08016	·27891	·42756	·54546	·64267	·72503	·85888	·96478	1·05192
60	·06009	·26160	·41193	·53099	·62904	·71206	·84687	·95344	1·04108
61	0·04005	0·24437	0·39641	0·51663	0·61553	0·69921	0·83498	0·94222	1·03037
62	·02002	·22721	·38098	·50237	·60213	·68647	·82320	·93112	1·01978
63	·00000	·21012	·36564	·48822	·58884	·67384	·81154	·92013	1·00930
64	—	·19309	·35039	·47416	·57566	·66132	·79998	·90925	0·99893
65	—	·17612	·33521	·46020	·56257	·64891	·78854	·89848	·98866
66	—	0·15919	0·32012	0·44632	0·54958	0·63659	0·77719	0·88782	0·97850
67	—	·14232	·30510	·43253	·53668	·62437	·76595	·87725	·96844
68	—	·12548	·29014	·41882	·52386	·61224	·75480	·86678	·95848
69	—	·10868	·27525	·40519	·51114	·60020	·74374	·85640	·94861
70	—	·09191	·26042	·39164	·49850	·58824	·73277	·84612	·93883
71	—	0·07516	0·24565	0·37816	0·48593	0·57637	0·72189	0·83592	0·92914
72	—	·05844	·23093	·36474	·47344	·56458	·71110	·82581	·91954
73	—	·04173	·21626	·35139	·46103	·55287	·70039	·81579	·91002
74	—	·02503	·20164	·33811	·44869	·54124	·68976	·80584	·90058
75	—	·00834	·18706	·32488	·43641	·52967	·67920	·79598	·89122
76	—	—	0·17252	0·31171	0·42420	0·51818	0·66872	0·78619	0·88194
77	—	—	·15802	·29859	·41205	·50676	·65831	·77648	·87274
78	—	—	·14355	·28553	·39997	·49540	·64798	·76684	·86361
79	—	—	·12911	·27251	·38794	·48410	·63771	·75727	·85455
80	—	—	·11470	·25954	·37596	·47287	·62751	·74777	·84556
81	—	—	0·10031	0·24661	0·36404	0·46169	0·61738	0·73833	0·83663
82	—	—	·08594	·23373	·35218	·45058	·60730	·72896	·82778
83	—	—	·07159	·22088	·34036	·43952	·59729	·71966	·81899
84	—	—	·05725	·20807	·32859	·42851	·58734	·71041	·81026
85	—	—	·04293	·19529	·31686	·41755	·57745	·70123	·80159
86	—	—	0·02862	0·18254	0·30518	0·40665	0·56761	0·69211	0·79298
87	—	—	·01431	·16983	·29354	·39579	·55783	·68304	·78443
88	—	—	·00000	·15714	·28194	·38498	·54810	·67403	·77594
89	—	—	—	·14448	·27038	·37421	·53842	·66507	·76750
90	—	—	—	·13184	·25885	·36349	·52879	·65617	·75912
91	—	—	—	0·11922	0·24736	0·35280	0·51922	0·64732	0·75079
92	—	—	—	·10662	·23590	·34216	·50968	·63852	·74252
93	—	—	—	·09404	·22447	·33156	·50020	·62976	·73429
94	—	—	—	·08147	·21307	·32099	·49076	·62106	·72611
95	—	—	—	·06891	·20170	·31046	·48136	·61240	·71798
96	—	—	—	0·05637	0·19035	0·29997	0·47201	0·60379	0·70990
97	—	—	—	·04383	·17903	·28951	·46269	·59522	·70186
98	—	—	—	·03130	·16773	·27907	·45342	·58670	·69387
99	—	—	—	·01878	·15645	·26867	·44419	·57822	·68593
100	—	—	—	·00626	·14520	·25830	·43499	·56978	·67802

Table 1 (*cont.*)

k \ n	225	250	300	350	400	k \ n	350	400
101	0·13396	0·24796	0·42583	0·56138	0·67016	**151**	0·17626	0·31517
102	·12274	·23764	·41670	·55302	·66234	**152**	·16900	·30860
103	·11153	·22735	·40761	·54470	·65456	**153**	·16174	·30203
104	·10034	·21708	·39856	·53641	·64682	**154**	·15450	·29548
105	·08916	·20683	·38953	·52817	·63912	**155**	·14726	·28895
106	0·07799	0·19661	0·38054	0·51996	0·63145	**156**	0·14003	0·28242
107	·06683	·18641	·37158	·51178	·62383	**157**	·13280	·27591
108	·05568	·17622	·36265	·50364	·61624	**158**	·12558	·26941
109	·04453	·16606	·35375	·49553	·60868	**159**	·11837	·26292
110	·03340	·15591	·34487	·48745	·60116	**160**	·11117	·25644
111	0·02226	0·14577	0·33602	0·47941	0·59367	**161**	0·10397	0·24998
112	·01113	·13566	·32720	·47139	·58622	**162**	·09678	·24352
113	·00000	·12555	·31841	·46341	·57880	**163**	·08959	·23707
114	—	·11546	·30963	·45545	·57141	**164**	·08240	·23064
115	—	·10538	·30089	·44753	·56405	**165**	·07522	·22421
116	—	0·09531	0·29216	0·43963	0·55672	**166**	0·06805	0·21779
117	—	·08526	·28346	·43176	·54942	**167**	·06088	·21138
118	—	·07520	·27478	·42392	·54215	**168**	·05371	·20498
119	—	·06516	·26612	·41610	·53491	**169**	·04654	·19859
120	—	·05513	·25748	·40831	·52770	**170**	·03938	·19220
121	—	0·04510	0·24885	0·40054	0·52051	**171**	0·03221	0·18583
122	—	·03507	·24025	·39280	·51335	**172**	·02505	·17946
123	—	·02505	·23167	·38508	·50622	**173**	·01789	·17310
124	—	·01503	·22310	·37738	·49911	**174**	·01074	·16674
125	—	·00501	·21455	·36970	·49203	**175**	·00358	·16040
126	—	—	0·20601	0·36205	0·48497	**176**	—	0·15406
127	—	—	·19749	·35442	·47794	**177**	—	·14772
128	—	—	·18898	·34681	·47093	**178**	—	·14139
129	—	—	·18049	·33922	·46394	**179**	—	·13507
130	—	—	·17201	·33164	·45698	**180**	—	·12875
131	—	—	0·16354	0·32409	0·45004	**181**	—	0·12244
132	—	—	·15508	·31656	·44312	**182**	—	·11613
133	—	—	·14664	·30904	·43622	**183**	—	·10983
134	—	—	·13820	·30154	·42934	**184**	—	·10353
135	—	—	·12978	·29406	·42248	**185**	—	·09723
136	—	—	0·12136	0·28659	0·41564	**186**	—	0·09094
137	—	—	·11296	·27914	·40883	**187**	—	·08465
138	—	—	·10456	·27171	·40203	**188**	—	·07837
139	—	—	·09617	·26429	·39524	**189**	—	·07209
140	—	—	·08778	·25689	·38848	**190**	—	·06581
141	—	—	0·07940	0·24950	0·38174	**191**	—	0·05954
142	—	—	·07103	·24212	·37501	**192**	—	·05326
143	—	—	·06266	·23475	·36830	**193**	—	·04699
144	—	—	·05430	·22740	·36160	**194**	—	·04072
145	—	—	·04594	·22006	·35492	**195**	—	·03445
146	—	—	0·03758	0·21274	0·34826	**196**	—	0·02819
147	—	—	·02923	·20542	·34161	**197**	—	·02192
148	—	—	·02088	·19812	·33498	**198**	—	·01566
149	—	—	·01252	·19082	·32836	**199**	—	·00939
150	—	—	·00417	·18354	·32176	**200**	—	·00313

Biometrika (1961), **48**, 1 and 2, *p.* 167
Printed in Great Britain

A distribution analogous to the Borel–Tanner

BY FRANK A. HAIGHT
Institute of Transportation and Traffic Engineering, University of California, Los Angeles

1. INTRODUCTION

Let $\alpha > 0$ and r be a positive integer. Then

$$p(x;r,\alpha) = A(x,r)\frac{\alpha^{x-r}}{(1+\alpha)^{2x-r}} \quad (x = r, r+1, \ldots)$$

is a probability distribution when

$$A(x,r) = \frac{r}{x}\binom{2x-r-1}{x-1}.$$

A proof of this fact for $r = 1$ will be found in a paper to be published shortly by the present writer (Haight, 1961); a slightly modified proof is valid for general r. In the theory of queues $p(x;r,\alpha)$ represents the probability that exactly x members of a queue will be served before the queue first vanishes, beginning with r members, and with α equal to the traffic intensity, assuming Poisson arrivals and negative exponential service periods. If the service periods are of constant length, the resulting distribution is that due to Borel (1942) and Tanner (1953). Also, if $r = 1$, $A(x,1)$ gives the number of ways brackets can be introduced into a product of x factors.

For any r, the generating function of $A(x,r)$ is

$$G(s) = \sum_x A(x,r)\,s^x = \left(\frac{2s}{1+\sqrt{(1-4s)}}\right)^r$$

and therefore the generating function of the probability distribution is

$$\phi(s) = \left(\frac{\alpha+1}{\alpha}\right)^r G\left(\frac{\alpha s}{(1+\alpha)^2}\right) = \left\{\frac{2s}{(1+\alpha)+\sqrt{\{(1+\alpha)^2-4\alpha s\}}}\right\}^r.$$

From this we find the $$\text{mean} = \frac{r}{1-\alpha},$$

which agrees with that of the Borel–Tanner distribution, and

$$\text{variance} = r\alpha\frac{1+\alpha}{(1-\alpha)^3}$$

which exceeds that of the Borel–Tanner distribution by a factor of $(1+\alpha)$. If $\alpha > 1$, $\phi(1) = \alpha^{-r}$, and therefore $p(\infty;r,\alpha) = 1-\alpha^{-r}$.

2. TABLES

In the table which follows, we give values of this distribution, correct to four places of decimals, in cumulative form, i.e.

$$P(x;r,\alpha) = \sum_{y=r}^{x} p(y;r,\alpha)$$

for the parameter values $r = 1$, $\alpha = 0{\cdot}01\ (0{\cdot}01)\ 0{\cdot}62$. In every case, the values of x taken are large enough to ensure that $P(x; r, \alpha) > 0{\cdot}999$. The upper limit of $0{\cdot}62$ for α was taken because this was the limit of the table of cumulative sums for the Borel–Tanner distribution published in a recent issue of this journal (Haight & Breuer, 1960). The function has not 'settled down' at this point and it is recognized that in practice it is likely that values of α just less than unity will occur. Plans are being made for additional computations in connexion with both functions, which will also include the case $r > 1$.

REFERENCES

BOREL, E. (1942). Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au problème de l'attente à un guichet. *C.R. Acad. Sci., Paris*, **214**, 452–6.

HAIGHT, FRANK A. (1961). Expected utility for queues servicing messages with exponentially decaying utility (to appear).

HAIGHT, FRANK A. & BREUER, MELVIN ALLEN (1960). The Borel–Tanner distribution. *Biometrika*, **47**, 143–50.

TANNER, J. C. (1953). A problem of interference between two queues. *Biometrika*, **40**, 58–69.

Table 1. *Cumulative probabilities* ($r = 1$)

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}01$	
1	0·99010
2	·99980
$\alpha = 0{\cdot}02$	
1	0·98039
2	·99923
$\alpha = 0{\cdot}03$	
1	0·97087
2	·99833
3	·99988
$\alpha = 0{\cdot}04$	
1	0·96154
2	·99710
3	·99973
$\alpha = 0{\cdot}05$	
1	0·95238
2	·99557
3	·99949
$\alpha = 0{\cdot}06$	
1	0·94340
2	·99378
3	·99916
$\alpha = 0{\cdot}07$	
1	0·93458
2	·99172
3	·99871
4	·99978
$\alpha = 0{\cdot}08$	
1	0·92593
2	·98944
3	·99815
4	·99964
$\alpha = 0{\cdot}09$	
1	0·91743
2	·98693
3	·99746
4	·99945

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}10$	
1	0·90909
2	·98422
3	·99664
4	·99921
$\alpha = 0{\cdot}11$	
1	0·90090
2	·98133
3	·99569
4	·99890
5	·99970
$\alpha = 0{\cdot}12$	
1	0·89286
2	·97827
3	·99461
4	·99852
5	·99956
$\alpha = 0{\cdot}13$	
1	0·88496
2	·97506
3	·99341
4	·99808
5	·99941
$\alpha = 0{\cdot}14$	
1	0·87719
2	·97169
3	·99205
4	·99753
5	·99918
$\alpha = 0{\cdot}15$	
1	0·86957
2	·96820
3	·99057
4	·99691
5	·99892
6	0·99961
$\alpha = 0{\cdot}16$	
1	0·86207
2	·96458
3	·98896
4	·99621
5	·99862

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}16$ (*cont.*)	
6	0·99948
$\alpha = 0{\cdot}17$	
1	0·85470
2	·96085
3	·98721
4	·99540
5	·99825
6	0·99931
$\alpha = 0{\cdot}18$	
1	0·84746
2	·95701
3	·98533
4	·99448
5	·99779
6	0·99907
$\alpha = 0{\cdot}19$	
1	0·84034
2	·95309
3	·98334
4	·99349
5	·99730
6	0·99883
7	·99948
$\alpha = 0{\cdot}20$	
1	0·83333
2	·94907
3	·98122
4	·99238
5	·99672
6	0·99853
7	·99932
$\alpha = 0{\cdot}21$	
1	0·82645
2	·94499
3	·97899
4	·99118
5	·99608
6	0·99819
7	·99914

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}22$	
1	0·81967
2	·94083
3	·97665
4	·98998
5	·99536
6	0·99779
7	·99892
8	·99946
$\alpha = 0{\cdot}23$	
1	0·81301
2	·93661
3	·97419
4	·98847
5	·99455
6	0·99732
7	·99865
8	·99930
$\alpha = 0{\cdot}24$	
1	0·80645
2	·93233
3	·97163
4	·98696
5	·99366
6	0·99680
7	·99834
8	·99906
$\alpha = 0{\cdot}25$	
1	0·80000
2	·92800
3	·96896
4	·98534
5	·99268
6	0·99620
7	·99797
8	·99889
9	·99928
$\alpha = 0{\cdot}26$	
1	0·79365
2	·92363
3	·96620
4	·98363
5	·99162

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}26$ (*cont.*)	
6	0·99555
7	·99757
8	·99864
9	·99917
$\alpha = 0{\cdot}27$	
1	0·78740
2	·91921
3	·96334
4	·98181
5	·99047
6	0·99482
7	·99711
8	·99835
9	·99904
$\alpha = 0{\cdot}28$	
1	0·78125
2	·91477
3	·96041
4	·97991
5	·98924
6	0·99402
7	·99659
8	·99802
9	·99883
10	·99930
$\alpha = 0{\cdot}29$	
1	0·77519
2	·91028
3	·95737
4	·97788
5	·98789
6	0·99312
7	·99599
8	·99761
9	·99855
10	·99911
$\alpha = 0{\cdot}30$	
1	0·76923
2	·90578
3	·95426
4	·97577
5	·98646

Table 1 (*cont.*)

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}30$ (*cont.*)	
6	0·99215
7	·99533
8	·99716
9	·99824
10	·99889
11	0·99929
$\alpha = 0{\cdot}31$	
1	0·76336
2	·90125
3	·95107
4	·97357
5	·98495
6	0·99112
7	·99462
8	·99668
9	·99792
10	·99868
11	0·99915
$\alpha = 0{\cdot}32$	
1	0·75758
2	·89671
3	·94781
4	·97127
5	·98333
6	0·98998
7	·99382
8	·99611
9	·99751
10	·99839
11	0·99895
12	·99931
$\alpha = 0{\cdot}33$	
1	0·75188
2	·89215
3	·94449
4	·96890
5	·98165
6	0·98879
7	·99297
8	·99551
9	·99709
10	·99809
11	0·99873
12	·99915
$\alpha = 0{\cdot}34$	
1	0·74627
2	·88758
3	·94109
4	·96642
5	·97985
6	0·98748
7	·99202
8	·99481
9	·99657
10	·99771
11	0·99845
12	·99894
13	·99927
$\alpha = 0{\cdot}35$	
1	0·74074
2	·88299
3	·93763
4	·96386
5	·97796
6	0·98609
7	·99099
8	·99405
9	·99601
10	·99729
11	0·99814
12	·99871
13	·99910
$\alpha = 0{\cdot}36$	
1	0·73529
2	·87841
3	·93412
4	·96123
5	·97600
6	0·98463
7	·98991
8	·99325
9	·99542
10	·99685
11	0·99781
12	·99847
13	0·99892
14	·99923
$\alpha = 0{\cdot}37$	
1	0·72993
2	·87382
3	·93055
4	·95851
5	·97394
6	0·98307
7	·98872
8	·99234
9	·99472
10	·99632
11	0·99741
12	·99816
13	·99868
14	·99905
$\alpha = 0{\cdot}38$	
1	0·72464
2	·86923
3	·92693
4	·95572
5	·97180
6	0·98142
7	·98746
8	·99138
9	·99398
10	·99575
11	0·99697
12	·99782
13	·99842
14	·99885
15	·99916
$\alpha = 0{\cdot}39$	
1	0·71942
2	·86464
3	·92326
4	·95284
5	·96956
6	0·97968
7	·98610
8	·99031
9	·99314
10	·99509
11	0·99645
12	·99741
13	·99809
14	·99858
15	·99894
16	0·99920
$\alpha = 0{\cdot}40$	
1	0·71429
2	·86006
3	·91956
4	·94992
5	·96727
6	0·97789
7	·98470
8	·98922
9	·99229
10	·99442
11	0·99592
12	·99699
13	·99777
14	·99834
15	·99876
16	0·99907
$\alpha = 0{\cdot}41$	
1	0·70922
2	·85548
3	·91581
4	·94691
5	·96487
6	0·97598
7	·98318
8	·98801
9	·99133
10	·99366
11	0·99532
12	·99652
13	·99739
14	·99803
15	·99851
16	0·99887
17	·99913
$\alpha = 0{\cdot}42$	
1	0·70423
2	·85091
3	0·91202
4	·94384
5	·96240
6	0·97400
7	·98159
8	·98673
9	·99030
10	·99283
11	0·99465
12	·99598
13	·99696
14	·99769
15	·99823
16	0·99864
17	·99895
18	·99919
$\alpha = 0{\cdot}43$	
1	0·69930
2	·84635
3	·90819
4	·94070
5	·95984
6	0·97192
7	·97990
8	·98535
9	·98917
10	·99190
11	0·99389
12	·99535
13	·99644
14	·99726
15	·99788
16	0·99835
17	·99871
18	·99900
$\alpha = 0{\cdot}44$	
1	0·69444
2	·84179
3	·90432
4	·93749
5	·95720
6	0·96975
7	·97812
8	·98389

Table 1 (*cont.*)

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}44$ (*cont.*)	
9	0·98797
10	·99091
11	0·99307
12	·99467
13	·99587
14	·99678
15	·99748
16	0·99802
17	·99843
18	·99875
19	·99900
$\alpha = 0{\cdot}45$	
1	0·68966
2	·83727
3	·90045
4	·93426
5	·95452
6	0·96753
7	·97628
8	·98237
9	·98671
10	·98987
11	0·99221
12	·99396
13	·99529
14	·99630
15	·99708
16	0·99769
17	·99816
18	·99853
19	·99882
20	·99905
$\alpha = 0{\cdot}46$	
1	0·68493
2	·83274
3	·89653
4	·93095
5	·95175
6	0·96521
7	·97434
8	·98074
9	·98535
10	·98873
11	0·99125
12	0·99315
13	·99460
14	·99572
15	·99659
16	0·99727
17	·99781
18	·99823
19	·99857
20	·99884
21	0·99905
$\alpha = 0{\cdot}47$	
1	0·68027
2	·82823
3	·89259
4	·92759
5	·94890
6	0·96281
7	·97231
8	·97903
9	·98390
10	·98750
11	0·99021
12	·99227
13	·99386
14	·99509
15	·99599
16	0·99675
17	·99735
18	·99783
19	·99821
20	·99852
21	0·99877
22	·99897
23	·99914
$\alpha = 0{\cdot}48$	
1	0·67568
2	·82375
3	·88865
4	·92420
5	·94602
6	0·96036
7	·97024
8	·97727
9	·98241
10	·98624
11	0·98914
12	·99136
13	·99308
14	·99443
15	·99549
16	0·99634
17	·99702
18	·99756
19	·99800
20	·99836
21	0·99865
22	·99889
23	·99909
$\alpha = 0{\cdot}49$	
1	0·67114
2	·81927
3	·88466
4	·92074
5	·94304
6	0·95780
7	·96804
8	·97539
9	·98079
10	·98485
11	0·98794
12	·99033
13	·99220
14	·99367
15	·99484
16	0·99577
17	·99652
18	·99713
19	·99763
20	·99803
21	0·99836
22	·99863
23	·99886
24	·99905
$\alpha = 0{\cdot}50$	
1	0·66667
2	·81482
3	·88066
4	·91724
5	·94000
6	0·95517
7	·96577
8	·97342
9	·97909
10	·98337
11	0·98665
12	·98921
13	·99122
14	·99282
15	·99410
16	0·99513
17	·99596
18	·99664
19	·99719
20	·99765
21	0·99803
22	·99834
23	·99860
24	·99881
25	·99899
26	0·99914
$\alpha = 0{\cdot}51$	
1	0·66225
2	·81038
3	·87664
4	·91369
5	·93690
6	0·95247
7	·96342
8	·97138
9	·97731
10	·98182
11	0·98531
12	·98804
13	·99020
14	·99193
15	·99332
16	0·99445
17	·99537
18	·99612
19	·99674
20	·99725
21	0·99768
22	·99804
23	·99834
24	0·99859
25	·99880
26	0·99898
27	·99913
$\alpha = 0{\cdot}52$	
1	0·65789
2	·80596
3	·87261
4	·91011
5	·93374
6	0·94970
7	·96099
8	·96925
9	·97544
10	·98018
11	0·98387
12	·98677
13	·98908
14	·99094
15	·99245
16	0·99368
17	·99469
18	·99552
19	·99621
20	·99678
21	0·99726
22	·99766
23	·99800
24	·99829
25	·99853
26	0·99874
27	·99891
28	·99906
$\alpha = 0{\cdot}53$	
1	0·65359
2	·80157
3	·86858
4	·90651
5	·93055
6	0·94688
7	·95850
8	·96705
9	·97350
10	·97847

Table 1 (*cont.*)

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}53$ (*cont.*)	
11	0·98236
12	·98544
13	·98791
14	·98991
15	·99154
16	0·99288
17	·99398
18	·99490
19	·99566
20	·99630
21	0·99684
22	·99729
23	·99767
24	·99800
25	·99828
26	0·99852
27	·99872
28	·99889
29	·99904
$\alpha = 0{\cdot}54$	
1	0·64935
2	·79720
3	·86453
4	·90285
5	·92728
6	0·94397
7	·95591
8	·96475
9	·97146
10	·97665
11	0·98073
12	·98398
13	·98660
14	·98873
15	·99048
16	0·99192
17	·99312
18	·99412
19	·99496
20	·99567
21	0·99627
22	·99678
23	·99721
24	·99758
25	·99790
26	0·99817
27	·99840
28	·99860
29	·99877
30	·99892
31	0·99905
$\alpha = 0{\cdot}55$	
1	0·64516
2	·79286
3	·86048
4	·89918
5	·92399
6	0·94103
7	·95329
8	·96241
9	·96937
10	·97479
11	0·97907
12	·98250
13	·98528
14	·98755
15	·98942
16	0·99098
17	·99228
18	·99337
19	·99429
20	·99507
21	0·99573
22	·99629
23	·99677
24	·99719
25	·99755
26	0·99786
27	·99813
28	·99836
29	·99856
30	·99873
31	0·99888
32	·99901
$\alpha = 0{\cdot}56$	
1	0·64103
2	·78854
3	·85643
4	·89548
5	·92064
6	0·93801
7	·95057
8	·95996
9	·96717
10	·97281
11	0·97729
12	·98090
13	·98384
14	·98626
15	·98826
16	0·98993
17	·99133
18	·99251
19	·99351
20	·99436
21	0·99509
22	·99572
23	·99626
24	·99673
25	·99713
26	0·99748
27	·99779
28	·99806
29	·99829
30	·99849
31	0·99867
32	·99883
33	·99897
34	·99910
$\alpha = 0{\cdot}57$	
1	0·63694
2	·78423
3	·85235
4	·89173
5	·91723
6	0·93492
7	·94778
8	·95744
9	·96489
10	·97075
11	0·97543
12	·97922
13	·98232
14	·98488
15	·98701
16	0·98880
17	·99031
18	·99159
19	·99268
20	·99361
21	0·99441
22	·99510
23	·99570
24	·99622
25	·99667
26	0·99706
27	·99740
28	·99770
29	·99796
30	·99819
31	0·99839
32	·99857
33	·99873
34	·99887
35	·99899
36	0·99910
$\alpha = 0{\cdot}58$	
1	0·63291
2	·77995
3	·84828
4	·88797
5	·91379
6	0·93179
7	·94493
8	·95485
9	·96253
10	·96860
11	0·97347
12	·97743
13	·98069
14	·98339
15	·98565
16	0·98755
17	·98916
18	·99053
19	·99170
20	·99271
21	0·99358
22	·99433
23	·99498
24	0·99555
25	·99605
26	0·99649
27	·99687
28	·99721
29	·99751
30	·99777
31	0·99800
32	·99820
33	·99838
34	·99854
35	·99868
36	0·99880
37	·99891
38	·99901
$\alpha = 0{\cdot}59$	
1	0·62893
2	·77571
3	·84422
4	·88419
5	·91031
6	0·92860
7	·94201
8	·95218
9	·96009
10	·96637
11	0·97143
12	·97557
13	·97899
14	·98184
15	·98423
16	0·98625
17	·98797
18	·98944
19	·99071
20	·99180
21	0·99275
22	·99358
23	·99430
24	·99493
25	·99548
26	0·99597
27	·99640
28	·99679
29	·99713
30	·99743

Table 1 (*cont.*)

x	$P(x; 1, \alpha)$
$\alpha = 0{\cdot}59$ (*cont.*)	
31	0·99769
32	·99793
33	·99814
34	·99833
35	·99850
36	0·99865
37	·99878
38	·99890
39	·99901
$\alpha = 0{\cdot}60$	
1	0·62500
2	·77149
3	·84016
4	·88040
5	·90681
6	0·92538
7	·93906
8	·94948
9	·95762
10	·96411
11	0·96936
12	·97367
13	·97724
14	·98023
15	·98275
16	0·98489
17	·98672
18	·98830
19	·98966
20	0·99084
21	0·99187
22	·99277
23	·99356
24	·99425
25	·99486
26	0·99540
27	·99588
28	·99630
29	·99668
30	·99701
31	0·99731
32	·99758
33	·99782
34	·99803
35	·99822
36	0·99839
37	·99855
38	·99869
39	·99882
40	·99893
41	0·99903
$\alpha = 0{\cdot}61$	
1	0·62112
2	·76729
3	·83609
4	·87656
5	·90323
6	0·92206
7	·93599
8	·94664
9	·95499
10	·96167
11	0·96710
12	·97158
13	·97531
14	·97844
15	·98109
16	0·98335
17	·98529
18	·98697
19	·98842
20	·98969
21	0·99080
22	·99177
23	·99262
24	·99337
25	·99404
26	0·99463
27	·99516
28	·99563
29	·99605
30	·99642
31	0·99675
32	·99705
33	·99732
34	·99756
35	·99778
36	0·99798
37	·99816
38	·99832
39	·99847
40	·99860
41	0·99872
42	·99883
43	·99893
44	·99902
$\alpha = 0{\cdot}62$	
1	0·61728
2	·76311
3	·83201
4	·87270
5	·89962
6	0·91870
7	·93286
8	·94373
9	·95229
10	·95917
11	0·96478
12	·96942
13	·97330
14	·97657
15	·97935
16	0·98173
17	·98378
18	·98556
19	0·98711
20	·98846
21	0·98965
22	·99070
23	·99162
24	·99244
25	·99317
26	0·99382
27	·99440
28	·99492
29	·99538
30	·99580
31	0·99617
32	·99651
33	·99681
34	·99708
35	·99733
36	0·99755
37	·99775
38	·99793
39	·99810
40	·99825
41	0·99839
42	·99852
43	·99864
44	·99875
45	·99885
46	0·99894
47	·99902

Biometrika (1961), **48**, 1 *and* 2, *p.* 175
Printed in Great Britain

Occupancy probability distribution critical points

BY W. L. NICHOLSON†
Operations Research and Synthesis Operation
Hanford Laboratories Operation

1. INTRODUCTION

Many statistical problems in the field of hypothesis testing involve a null hypothesis probability law for the random independent allocation of N identical balls in K equally likely cells. The alternatives to the equally likely null hypothesis are unequal cell probabilities and/or uniform systematic allocation. When the number of balls N is reasonably large, say $N \geqslant 5K$, the classical χ^2 goodness-of-fit test criterion can be used to decide among the null hypothesis and the two alternatives, a small χ^2 value favouring the systematic alternative and a large χ^2 value favouring the unequal cell probabilities alternative. If N is small, and if $K > 2$, neither χ^2 nor the well-tabulated binomial distribution is applicable as a test criterion.

Stevens (1937) and later David (1950) suggest as a test criterion the number X of 'occupied' cells, or equivalently the number $K - X$ of 'empty' cells. Too many empty cells, a significantly small value of X, favours an unequal cell probabilities alternative, while too few empty cells, a significantly large value of X, favours a systematic smooth allocation of the balls within the K cells. For $N = K$, David indicates that the test criterion X has good power against both alternatives. As a result of numerical checks she suggests the use of the exact probability distribution of X for $N = K \leqslant 20$, a normal approximation for $20 < N < 30$ with $K = 25$ in all cases, and for $N \geqslant 30$ a χ^2 test based on expected cell frequencies of at least 5. At the cross-over point of $N = 30$ she shows with an example that the occupancy criterion has approximately the same power as that of the χ^2 goodness-of-fit criterion for a test concerning the standard deviation of a normal distribution with known mean. For the case of $N = K$, David tables the probability distribution of $K - X$ and gives a short table of suggested critical points.

In the goodness-of-fit type problems considered by David a completely specified probability distribution is partitioned into a K-cell histogram with each cell having probability $1/K$ under the null hypothesis. The value of K can always be selected equal to the sample size N. An example is provided by Thomas (1951) in a null hypothesis stating that K Poisson distributions all have the same but unknown mean where, in general, it is not possible to choose $K = N$. Here, the conditional joint distribution of the K Poisson variables given their sum, say N, is multinomial with equally likely cells. Thomas is only concerned with large sample theory and uses appropriate normal distribution tests. There are, however, many applications of this technique (for example, in the field of low level radioactive counting) where a small sample size prohibits the use of the normal distribution. It is primarily for these cases that David's tables have been extended.

† General Electric Company, Richland, Washington. Work performed under Contract No. AT(45-1). 1350 for the U.S. Atomic Energy Commission.

2. DESCRIPTION OF THE TABLE

The occupancy probability distribution function $H(X; K, N)$, the probability of having exactly X occupied cells when N balls are randomly and independently distributed among K equally probable cells, can be shown to be (see David, 1950 or Feller, 1950)

$$H(X; K, N) = \binom{K}{X} \sum_{j=0}^{X} (-1)^j \binom{X}{j} \left(\frac{X-j}{K}\right)^N.$$

A table of values of the occupancy probability distribution function $H(X; K, N)$ for all meaningful combinations of $N = 1\,(1)\,30$, $K = 2\,(1)\,20$, and $X = 1\,(1)\,K$ has been calculated on IBM 650 of the Hanford Atomic Products Operation using the H-function recursion formula,

$$H(X; K, N) = \frac{K+1-X}{K} H(X-1; K, N-1) + \frac{X}{K} H(X; K, N-1).$$

The recursion formula is a probabilistic statement of the fact that (thinking of the balls as being dropped sequentially into the K cells) X occupied cells after the Nth ball is dropped implies either $X-1$ occupied cells after the $(N-1)$st ball is dropped with the Nth ball falling in an empty cell, or X occupied cells after the $(N-1)$st ball is dropped with the Nth ball falling into an occupied cell. The boundary conditions for the recursion relationship are

$$H(1; K, N) = K^{1-N}, \quad H(K; K, N) = K^{1-N}(N-1)!/(N-K)!.$$

The table, which is a direct extension of David's for $K = N = 3\,(1)\,20$, appears in Nicholson (1960), copies of which are available from the author.

To facilitate the performing of hypothesis tests using the occupancy criterion X, critical points corresponding to upper and lower tail areas of 0·01, 0·025, 0·05 and 0·10 were computed as appropriate sums of the entries in the above-mentioned table and appear below in tabular form. For each K, X and lower tail area α the smallest value of N, say N_1, which satisfies

$$\alpha_1 = \sum_{j=1}^{X} H(j; K, N) \leqslant \alpha,$$

is tabled in bold type along with the corresponding α_1, set in smaller type. For any $N < N_1$ the tail area at and below X is $> \alpha$. A result of $\leqslant X$ occupied cells would not be significant at the 100α % level, and there would be no reason to reject the null hypothesis in favour of an unequal cell probability alternative. For $N \geqslant N_1$ the tail area at and below X is $\leqslant \alpha_1$. A result of $\leqslant X$ occupied cells would therefore be significant at the 100α % level, and rejection of the null hypothesis in favour of an unequal cell probability alternative would be in order.

Similarly, for each K, X and upper tail area α the largest value of N, say N_2, which satisfies

$$\alpha_2 = \sum_{j=X}^{K} H(j; K, N) \leqslant \alpha,$$

is tabled along with the corresponding α_2. For any $N > N_2$ the tail area at and above X is $> \alpha$, so a result of $\geqslant X$ occupied cells would not be significant at the 100α % level, and there would be no reason to reject the null hypothesis in favour of a systematic smooth allocation alternative. For $N \leqslant N_2$ the tail area at and above X is $\leqslant \alpha_2$, so a result of $\geqslant X$ occupied cells would be significant at the 100α % level, and rejection of the null hypothesis in favour of the smooth allocation alternative would be in order.

For each K the lower tail table lists all X values for which the 10 % level of significance N_1 entry is $\leqslant 30$. For each K the upper tail table starts with the smallest X value for which a 10 % level of significance N_2 entry exists and lists all subsequent X values for which the 1 % level of significance N_2 entry is $\leqslant 30$. Dash marks indicate the non-existence of N_2 entries. In both lower and upper tail tables the entry > 30 is used in lieu of the exact value which is beyond the range of calculation from the occupancy probability distribution table used to compute the critical points.

To illustrate the use of the table of critical points consider a test based on $K = 10$ cells which results in $X = 6$ occupied cells. The minimum number of balls which allows rejection of the null hypothesis in favour of an unequal cell probabilities alternative is $N_1 = 15$ at the 10 % level of significance (the exact probability of not more than 6 occupied cells when 15 balls are distributed randomly and independently in 10 cells is $\alpha_1 = 0{\cdot}070$), $N_1 = 16$ at the 5 % level of significance, $N_1 = 18$ at the 2·5 % level of significance, and $N_1 = 20$ at the 1 % level of significance. With $K = 10$ cells and $X = 6$ occupied cells rejection of the null hypothesis in favour of a systematic smooth alternative is not possible at any significance level $\leqslant 10$ %. In order to get significance more cells must be occupied. For example, with $K = 10$ cells and $X = 8$ occupied cells the maximum number of balls which allows rejection of the null hypothesis in favour of a systematic smooth alternative is $N_1 = 9$ at the 10 % level of significance and $N_1 = 8$ at the 2·5 % level of significance. With $K = 10$ rejection of the null hypothesis in favour of the systematic alternative at the 1 % level of significance demands at least $X = 9$ cells occupied, specifically, with $X = 9$ necessarily $N = N_1 = 9$, and with $X = 10$ necessarily $10 \leqslant N \leqslant N_1 = 12$.

The author is indebted to several members of the Hanford Atomic Products Operation for assistance in the calculation of the tables. W. F. Stevenson wrote the IBM 650 program and compiled the machine output for the occupancy probability distribution values. P. E. Leaverton checked the machine calculations and computed the critical points listed in the table below. Operating personnel of the Hanford Atomic Products Electronic Data Processing Operation handled the technical details of the IBM calculation. Appreciation is also expressed to E. S. Pearson for several suggestions which simplify the presentation of the critical points table.

References

David, F. N. (1950). Two combinatorial tests of whether a sample has come from a given population. *Biometrika*, **37**, 97–110.

Feller, W. (1950). *Probability Theory and Its Applications*. New York: John Wiley and Sons, Inc.

Nicholson, W. L. (1960). *Occupancy Probability Tables Based on the Multinomial Distribution for Equally Probable Events*. Report HW-57502 REV.

Stevens, W. L. (1937). Significance of grouping. *Ann. Eugen., Lond.*, **8**, 57–69.

Thomas, M. (1951). Some tests for randomness in plant populations. *Biometrika*, **38**, 102–11.

Table 1. *Critical points for occupancy probability distributions*

Lower tail tests. Alternative: unequal cell probabilities

	X	Lower tail probability level α							
		0·01		0·025		0·05		0·10	
		N_1	α_1	N_1	α_1	N_1	α_1	N_1	α_1
$K=2$	1	8	·008	7	·016	6	·031	5	·062+
$K=3$	1	6	·004	5	·012	4	·037	4	·037
	2	15	·007	12	·023	11	·035	9	·078
$K=4$	1	5	·004	4	·016	4	·016	3	·062+
	2	10	·006	8	·023	7	·046	6	·092
	3	21	·010−	18	·023	16	·040	13	·094
$K=5$	1	4	·008	4	·008	3	·040	3	·040
	2	8	·007	7	·016	6	·040	5	·098
	3	14	·008	12	·021	11	·035	9	·096
	4	28	·010−	24	·024	21	·050−	18	·089
$K=6$	1	4	·005	4	·005	3	·028	3	·028
	2	7	·007	6	·020	6	·020	5	·059
	3	11	·010−	10	·019	9	·037	8	·071
	4	18	·010−	16	·022	14	·049	13	·072
	5	> 30	—	> 30	—	27	·044	23	·089
$K=7$	1	4	·003	4	·003	3	·020	3	·020
	2	7	·003	6	·011	5	·038	5	·038
	3	10	·007	9	·016	8	·036	7	·080
	4	15	·008	13	·023	12	·038	11	·065
	5	23	·009	20	·024	18	·046	16	·088
	6	> 30	—	> 30	—	> 30	—	28	·092
$K=8$	1	4	·002	3	·016	3	·016	3	·016
	2	6	·007	6	·007	5	·026	4	·098
	3	9	·008	8	·020	7	·050−	7	·050−
	4	13	·008	12	·015	11	·030	10	·056
	5	19	·007	17	·017	15	·042	13	·100−
	6	28	·009	25	·020	22	·046	20	·080
$K=9$	1	4	·001	3	·012	3	·012	3	·012
	2	6	·004	5	·018	5	·018	4	·078
	3	9	·004	8	·012	7	·033	6	·090
	4	12	·007	11	·015	10	·031	9	·065
	5	16	·009	15	·016	13	·048	12	·081
	6	23	·007	20	·022	18	·048	16	·099
	7	> 30	—	29	·023	26	·048	23	·097
$K=10$	1	3	·010	3	·010	3	·010	2	·100
	2	6	·003	5	·014+	5	·014+	4	·064
	3	8	·007	7	·022	7	·022	6	·068
	4	11	·008	10	·018	9	·041	8	·093
	5	15	·007	13	·024	12	·045	11	·082
	6	20	·008	18	·018	16	·045	15	·070
	7	27	·007	24	·020	22	·039	20	·074
	8	> 30	—	> 30	—	> 30	—	27	·094
$K=11$	1	3	·008	3	·008	3	·008	2	·091
	2	6	·002	5	·010+	5	·010+	4	·053

Random independent allocation of N balls into K cells so that $\leqslant X$ cells are occupied. For given K and X and any $N \geqslant N_1$ the tail area at and below X is $\leqslant \alpha_1$.

Upper tail tests. Alternative: systematic allocation

	X	Upper tail probability level α							
		0·01		0·025		0·05		0·10	
		N_2	α_2	N_2	α_2	N_2	α_2	N_2	α_2
$K=2$		—	—	—	—	—	—	—	—
$K=3$		—	—	—	—	—	—	—	—
$K=4$	4	—	—	—	—	—	—	4	·094
$K=5$	5	—	—	—	—	5	·038	5	·038
$K=6$	5	—	—	—	—	—	—	5	·093
	6	—	—	6	·015	6	·015	7	·054
$K=7$	6	—	—	—	—	6	·043	6	·043
	7	7	·006	8	·024	8	·024	9	·058
$K=8$	6	—	—	—	—	—	—	6	·077
	7	—	—	7	·019	7	·019	8	·067
	8	8	·002	9	·011	10	·028	12	·093
$K=9$	7	—	—	—	—	7	·038	7	·038
	8	8	·008	8	·008	9	·035	10	·083
	9	10	·005	11	·013	12	·029	14	·081
$K=10$	7	—	—	—	—	—	—	7	·060
	8	—	—	8	·018	8	·018	9	·069
	9	9	·004	10	·017	11	·044	12	·087
	10	12	·006	13	·014	15	·046	16	·070
$K=11$	7	—	—	—	—	—	—	7	·085
	8	—	—	—	—	8	·031	8	·031

Random independent allocation of N balls into K cells so that $\geqslant X$ cells are occupied. For given K and X and any $N \leqslant N_2$ the tail area at and above X is $\leqslant \alpha_2$.

Lower tail tests. Alternative: unequal cell probabilities

Upper tail tests. Alternative: systematic allocation

	X	Lower tail probability level α								X	Upper tail probability level α							
		0·01		0·025		0·05		0·10			0·01		0·025		0·05		0·10	
		N_1	α_1	N_1	α_1	N_1	α_1	N_1	α_1		N_2	α_2	N_2	α_2	N_2	α_2	N_2	α_2
$K=11$	3	8	·004	7	·016	7	·016	6	·052	9	9	·008	9	·008	10	·036	11	·089
	4	11	·004	10	·011	9	·027	8	·073	10	11	·008	12	·023	13	·048	14	·086
	5	14	·006	13	·013	12	·026	11	·052	11	14	·007	15	·014	17	·040	19	·084
	6	18	·007	16	·021	15	·036	14	·062									
	7	23	·009	21	·020	19	·045	17	·097									
	8	> 30	—	28	·019	25	·046	23	·081									
$K=12$	1	3	·007	3	·007	3	·007	2	·083	8	—	—	—	—	8	·046	8	·046
	2	5	·008	5	·008	4	·045	4	·045	9	—	—	9	·016	9	·016	10	·062
	3	8	·003	7	·011	6	·041	6	·041	10	10	·004	11	·018	12	·049	13	·099
	4	10	·007	9	·019	9	·019	8	·050+	11	12	·004	13	·011	15	·048	16	·082
	5	13	·007	12	·016	11	·034	10	·072	12	16	·007	18	·023	19	·035	22	·094
	6	17	·006	15	·020	14	·036	13	·065									
	7	21	·008	19	·020	18	·033	16	·081									
	8	27	·007	24	·022	22	·046	20	·091									
	9	> 30	—	> 30	—	29	·042	26	·089									
$K=13$	1	3	·006	3	·006	3	·006	2	·077	8	—	—	—	—	—	—	8	·064
	2	5	·006	5	·006	4	·039	4	·039	9	—	—	9	·025-	9	·025	10	·092
	3	7	·008	7	·008	6	·033	6	·033	10	10	·008	10	·008	11	·034	12	·085
	4	10	·004	9	·013	8	·038	8	·038	11	12	·009	12	·009	13	·027	14	·058
	5	12	·009	11	·022	11	·022	10	·052	12	14	·006	15	·014	17	·048	18	·077
	6	16	·005	14	·022	13	·042	12	·080	13	18	·007	20	·020	22	·044	24	·079
	7	19	·009	18	·016	16	·048	15	·080									
	8	24	·008	22	·020	20	·047	19	·070									
	9	30	·009	28	·018	25	·048	23	·090									
	10	> 30	—	> 30	—	> 30	—	29	·010-									
$K=14$	1	3	·005	3	·005	3	·005	2	·071	8	—	—	—	—	—	—	8	·082
	2	5	·005	5	·005	4	·034	4	·034	9	—	—	—	—	9	·035	9	·035
	3	7	·006	7	·006	6	·027	6	·027	10	—	—	10	·013	10	·013	11	·053
	4	9	·009	9	·009	8	·029	7	·086	11	11	·004	12	·018	13	·049	14	·100-
	5	12	·006	11	·016	10	·038	9	·090	12	13	·004	14	·014	15	·032	16	·062
	6	15	·006	14	·013	13	·028	12	·057	13	16	·007	17	·015	19	·046	20	·070
	7	18	·009	17	·016	16	·029	14	·090	14	20	·007	22	·018	24	·037	27	·084
	8	22	·009	20	·024	19	·039	17	·097									
	9	27	·009	25	·021	23	·044	21	·091									
	10	> 30	—	> 30	—	29	·039	26	·090									
$K=15$	1	3	·004	3	·004	3	·004	2	·067	9	—	—	—	—	9	·047	9	·047
	2	5	·004	5	·004	4	·029	4	·029	10	—	—	10	·019	10	·019	11	·076
	3	7	·005	6	·022	6	·022	5	·094	11	11	·006	11	·006	12	·029	13	·077
	4	9	·007	8	·023	8	·023	7	·072	12	13	·009	13	·009	14	·027	15	·060
	5	12	·004	11	·011	10	·029	9	·072	13	15	·007	16	·018	17	·036	18	·063
	6	14	·008	13	·019	12	·041	11	·086	14	18	·008	19	·015	21	·043	23	·090
	7	17	·009	16	·018	15	·034	14	·063	15	23	·010-	25	·022	27	·042	30	·088
	8	21	·007	19	·022	18	·038	17	·063									
	9	25	·009	23	·022	22	·034	20	·079									
	10	> 30	—	28	·022	26	·044	24	·084									
	11	> 30	—	> 30	—	> 30	—	29	·092									
$K=16$	1	3	·004	3	·004	3	·004	2	·062+	9	—	—	—	—	—	—	9	·060
	2	5	·003	5	·003	4	·026	4	·026	10	—	—	—	—	10	·026	10	·026
	3	7	·004	6	·018	6	·018	5	·084	11	11	·010-	11	·010-	12	·044	12	·044
	4	9	·005	8	·018	8	·018	7	·061	12	12	·003	13	·016	14	·045	15	·095
	5	11	·008	10	·022	10	·022	9	·058	13	14	·005	15	·015	16	·035	17	·067
	6	14	·005	13	·013	12	·030	11	·067	14	17	·009	18	·020	19	·038	21	·096
	7	17	·005	15	·023	14	·045	13	·086	15	20	·008	21	·015	23	·039	25	·080
	8	20	·007	18	·023	17	·041	16	·071	16	25	·009	27	·019	30	·046	> 30	—
	9	24	·007	22	·019	20	·048	19	·076									

Table 1 (*cont.*)

Lower tail tests. Alternative: unequal cell probabilities

Upper tail tests. Alternative: systematic allocation

	X	Lower tail probability level α								X	Upper tail probability level α							
		0·01		0·025		0·05		0·10			0·01		0·025		0·05		0·10	
		N_1	α_1	N_1	α_1	N_1	α_1	N_1	α_1		N_2	α_2	N_2	α_2	N_2	α_2	N_2	α_2
$K=16$	10	28	·009	26	·021	24	·047	22	·097									
	11	> 30	—	> 30	—	29	·044	27	·081									
$K=17$	1	3	·003	3	·003	3	·003	2	·059	9	—	—	—	—	—	—	9	·074
	2	5	·003	4	·023	4	·023	4	·023	10	—	—	—	—	10	·035	10	·035
	3	7	·003	6	·016	6	·016	5	·072	11	—	—	11	·014	11	·014	12	·061
	4	9	·004	8	·014	8	·014	7	·052	12	12	·005	13	·025⁻	13	·025	14	·068
	5	11	·006	10	·017	9	·047	9	·047	13	14	·008	14	·008	15	·026	16	·058
	6	13	·009	12	·022	12	·022	11	·052	14	16	·008	17	·020	18	·040	19	·072
	7	16	·007	15	·015	14	·032	13	·066	15	18	·005	20	·022	21	·038	23	·090
	8	19	·008	18	·015	17	·027	15	·090	16	22	·008	24	·023	25	·035	28	·093
	9	23	·006	21	·018	19	·050⁻	18	·081	17	27	·010⁻	30	·022	30	·022	> 30	—
	10	27	·007	25	·017	23	·040	21	·090									
	11	> 30	—	29	·021	27	·043	25	·085									
	12	> 30	—	> 30	—	> 30	—	30	·079									
$K=18$	1	3	·003	3	·003	3	·003	2	·056	9	—	—	—	—	—	—	9	·089
	2	5	·002	4	·021	4	·021	4	·021	10	—	—	—	—	10	·044	10	·044
	3	7	·002	6	·013	6	·013	5	·067	11	—	—	11	·020	11	·020	12	·080
	4	9	·003	8	·012	7	·044	7	·044	12	12	·008	12	·008	13	·036	14	·094
	5	11	·004	10	·013	9	·039	9	·039	13	13	·003	14	·014	15	·040	16	·087
	6	13	·006	12	·015	11	·041	10	·098	14	15	·004	16	·014	17	·035	18	·068
	7	16	·005	14	·023	14	·023	13	·050⁺	15	17	·004	19	·023	20	·044	21	·073
	8	18	·009	17	·018	16	·036	15	·072	16	20	·006	22	·022	23	·037	25	·084
	9	22	·006	20	·019	19	·033	17	·095	17	24	·008	26	·021	28	·045	30	·082
	10	25	·009	23	·024	22	·038	20	·093	18	30	·010⁻	> 30	—	> 30	—	> 30	—
	11	30	·007	27	·023	26	·035	24	·074									
	12	> 30	—	> 30	—	30	·035	28	·078									
$K=19$	1	3	·003	3	·003	3	·003	2	·053	10	—	—	—	—	—	—	10	·055
	2	5	·002	4	·019	4	·019	4	·019	11	—	—	—	—	11	·026	11	·026
	3	7	·002	6	·011	6	·011	5	·061	12	—	—	12	·011	13	·049	13	·049
	4	8	·010⁻	8	·010	7	·037	7	·037	13	13	·004	14	·021	14	·021	15	·058
	5	11	·003	10	·010⁺	9	·032	8	·096	14	15	·007	16	·023	16	·023	17	·054
	6	13	·005	12	·013	11	·033	10	·083	15	17	·008	18	·020	19	·041	20	·074
	7	15	·007	14	·017	13	·039	12	·085	16	19	·006	20	·013	22	·046	23	·073
	8	18	·006	17	·013	16	·026	14	·097	17	22	·007	24	·022	25	·036	27	·077
	9	21	·006	19	·022	18	·035	17	·071	18	26	·008	28	·019	30	·040	> 30	—
	10	24	·008	22	·024	21	·042	19	·065									
	11	28	·008	26	·020	24	·047	23	·081									
	12	> 30	—	30	·021	28	·038	26	·090									
	13	> 30	—	> 30	—	> 30	—	30	·097									
$K=20$	1	3	·002⁺	3	·002⁺	2	·050	2	·050	10	—	—	—	—	—	—	10	·066
	2	5	·002	4	·017	4	·017	4	·017	11	—	—	—	—	11	·033	11	·033
	3	6	·010⁻	6	·010	6	·010	5	·055	12	—	—	12	·015	12	·015	13	·063
	4	8	·008	8	·008	7	·033	7	·033	13	13	·006	13	·006	14	·029	15	·078
	5	10	·008	10	·008	9	·027	8	·084	14	14	·002	15	·011	16	·035	17	·078
	6	12	·010⁻	12	·010	11	·027	10	·070	15	16	·004	17	·013	18	·033	19	·065
	7	15	·005	14	·013	13	·031	12	·070	16	18	·004	20	·025⁻	21	·047	22	·078
	8	17	·004	16	·019	15	·039	14	·078	17	21	·007	22	·015	24	·046	25	·071
	9	20	·008	19	·015	18	·029	16	·095	18	24	·007	26	·022	27	·034	30	·095
	10	23	·009	22	·015	20	·046	19	·076	19	28	·007	30	·018	> 30	—	> 30	—
	11	27	·007	25	·019	23	·047	22	·074									
	12	> 30	—	29	·017	27	·039	25	·084									
	13	> 30	—	> 30	—	> 30	—	29	·079									

Random independent allocation of N balls into K cells so that $\leqslant X$ cells are occupied. For given K and X and any $N \geqslant N_1$ the tail area at and below X is $\leqslant \alpha_1$.

Random independent allocation of N balls into K cells so that $\geqslant X$ cells are occupied. For given K and X and any $N \leqslant N_2$ the tail area at and above X is $\leqslant \alpha_2$.

Biometrika (1961), **48**, 1 *and* 2, *p.* 181
Printed in Great Britain

Test of independence in intraclass 2×2 tables

BY MASASHI OKAMOTO AND GORO ISHII
Osaka University and Atomic Bomb Casualty Commission, Hiroshima, Japan

1. INTRODUCTION

If for each member of a sample of size n a pair (A_i, A_j) of characteristics A_i $(i = 1, 2, \ldots, k)$ is observed and if the pair (A_i, A_j) is indistinguishable from the reversed pair (A_j, A_i), the data can be represented by Table 1, where n_{ij} $(i \leqslant j)$ denotes the frequency of the pair (A_i, A_j) and $n = \sum\sum_{i \leqslant j} n_{ij}$.

Table 1. *Pair table*

	A_1	A_2	...	A_k	Total
A_1	n_{11}	n_{12}	...	n_{1k}	
A_2		n_{22}	...	n_{2k}	
⋮			⋱	⋮	
A_k				n_{kk}	
Total					n

Table 2. *Component table*

	A_1	A_2	...	A_k	Total
A_1	$2n_{11}$	n_{12}	...	n_{1k}	n_1
A_2	n_{21}	$2n_{22}$	...	n_{2k}	n_2
⋮	⋮	⋮	⋱	⋮	⋮
A_k	n_{k1}	n_{k2}	...	$2n_{kk}$	n_k
Total	n_1	n_2	...	n_k	$2n$

Such a situation is met with in some fields of biology, for example, in classifying several pairs of twins with regard to certain attributes of an individual. Another example is the M, N blood groups dealt with in medical research.

If we now consider for each i the number of components of the pairs having the characteristic A_i, we obtain Table 2, where we have put $n_{ji} = n_{ij}$ and $n_i = 2n_{ii} + \sum_{j \neq i} n_{ij}$. Tables 1 and 2 correspond one-to-one to each other and we shall refer to the former as the pair table and to the latter as the component table. Moreover, we shall call either of them an intraclass contingency table in contrast to an ordinary contingency table.

As for an ordinary contingency table, we are often interested in the problem of independence for our intraclass case. If we denote the probability of occurrence of the pair (A_i, A_j) by p_{ij} and that of A_i by p_i, then the hypothesis of independence is represented by

$$H_0: \quad p_{ii} = p_i^2, \quad p_{ij} = 2p_i p_j \quad (i < j). \tag{1}$$

When as a special case the number k of characteristics A_i equals two, A and $\bar{A}$, there arise three types of pair, (A, A), $(A, \bar{A})$ and $(\bar{A}, \bar{A})$; hence Tables 1 and 2 reduce to Tables 3 and 4, respectively, either of which we call an intraclass 2×2 (contingency) table. Let p_{AA}, $p_{A\bar{A}}$, $p_{\bar{A}\bar{A}}$, p_A and $p_{\bar{A}}$ be the probabilities of the pairs (A, A), $(A, \bar{A})$, $(\bar{A}, \bar{A})$, the components A and $\bar{A}$, respectively. Then the hypothesis of independence reduces to

$$H_0: \quad p_{AA} = p_A^2, \quad p_{A\bar{A}} = 2p_A p_{\bar{A}}, \quad p_{\bar{A}\bar{A}} = p_{\bar{A}} \tag{2}$$

which shows that the test of independence is identical with the test of fit to the binomial distribution $B(2, p_A)$.

Table 3. *Pair table*

	A	$\bar{A}$	Total
A	a	b	.
$\bar{A}$	.	c	.
Total	.	.	n

Table 4. *Component table*

	A	$\bar{A}$	Total
A	$2a$	b	r
$\bar{A}$	b	$2c$	s
Total	r	s	$2n$

2. INTRACLASS CONTINGENCY TABLE

The probability of observing Table 1 in a random sampling with only the total n fixed is given by the multinomial distribution

$$P(n_{ii}, n_{ij}) = \frac{n!}{\prod_i n_{ii}! \prod_{i<j}\prod n_{ij}!} \left(\prod_i p_{ii}^{n_{ii}}\right)\left(\prod_{i<j}\prod p_{ij}^{n_{ij}}\right), \tag{3}$$

which, under the hypothesis (1) of independence, becomes

$$P(n_{ii}, n_{ij}) = \frac{n!}{\prod_i n_{ii}! \prod_{i<j}\prod n_{ij}!} 2^{\sum\sum_{i<j} n_{ij}} \prod_i p_i^{n_i}, \tag{4}$$

where $n_{ji} = n_{ij}$ and $n_i = 2n_{ii} + \sum_{j+i} n_{ij}$ as before.

Since (3) is a general term of the multinomial expansion of

$$\left(\sum_i p_{ii} + \sum_{i<j}\sum p_{ij}\right)^n$$

which reduces under the hypothesis (1) to

$$\left(\sum_i p_i^2 + \sum_{i<j}\sum 2p_i p_j\right)^n \equiv \left(\sum_i p_i\right)^{2n},$$

and since the probability $P(n_i)$ of the 'marginal' frequencies n_i $(i = 1, 2, \ldots, k)$ is given by the summation of (4) over the set of $\{n_{ii}, n_{ij}\}$ satisfying $2n_{ii} + \sum_{j+i} n_{ij} = n_i$, we obtain from a general term of the expansion of $\left(\sum_i p_i\right)^{2n}$ the formula

$$P(n_i) = \frac{(2n)!}{\prod_i n_i!} \prod_i p_i^{n_i}. \tag{5}$$

Dividing (4) by (5), we get the conditional probability under H_0 of Table 1 when the n_i's are given

$$P(n_{ii}, n_{ij} | n_i) = \frac{n! \prod_i n_i!}{(2n)! \prod_i n_{ii}! \prod_{i<j}\prod n_{ij}!} 2^{\sum\sum_{i<j} n_{ij}}. \tag{6}$$

From the probability distribution (6) we can calculate easily the factorial moments up to the second order

$$\left.\begin{aligned} E(n_{ii}) &= \frac{n_i^{(2)}}{2(2n-1)}, \quad E(n_{ii}^{(2)}) = \frac{n_i^{(4)}}{4(2n-1)(2n-3)}, \\ E(n_{ij}) &= \frac{n_i n_j}{2n-1}, \quad E(n_{ij}^{(2)}) = \frac{n_i^{(2)} n_j^{(2)}}{(2n-1)(2n-3)}, \end{aligned}\right\} \quad (i \neq j), \tag{7}$$

where $n^{(2)}$ stands for $n(n-1)$ and $n^{(4)}$ for $n(n-1)(n-2)(n-3)$. For the statistic

$$\chi^2 = \sum_{i<j}\sum \frac{[n_{ij} - E(n_{ij})]^2}{E(n_{ij})}, \tag{8}$$

which may be used for testing the hypothesis (1), we have in view of (7)

$$E(\chi^2) = \frac{1}{2}\left(1 + \frac{1}{2n-3}\right) k(k-1). \tag{9}$$

This corresponds to Haldane's (1939) result $E(\chi^2) = \{1 + 1/(n-1)\}(r-1)(s-1)$ for the ordinary contingency table. The calculation for $V(\chi^2)$ seems to be even more tedious than that of Haldane's case, which is itself considerably complicated and therefore we have not carried it through. On the other hand it is seen by the usual large sample theory that under the unconditional distribution (4) the statistic χ^2 follows in the limit a χ^2 distribution with $\frac{1}{2}k(k-1)$ degrees of freedom as $n \to \infty$. One of the authors (Okamoto, 1959) proved that this is the case also with the conditional distribution (6).

3. Conditional test for an intraclass 2 × 2 table

We shall consider the test of the hypothesis H_0 of independence given by (2) for an intraclass 2×2 table represented in the form of Table 3 or 4. The χ^2 test based on the statistic (8) may be used if the expected values of the frequencies a, b and c under the null hypothesis are all moderately large. But if the expected values are not so large we recommend an exact test, the intraclass analogue of Fisher's exact test applied to the ordinary 2×2 contingency table. This test was introduced by Fisher in 1934 into the fifth edition of *Statistical Methods for Research Workers* (see also, Yates, 1934; Fisher, 1935), and has led to much discussion among mathematical statisticians, for instance, Barnard (1947), Pearson (1947, 1955) and Fisher (1955).

As a special case of the equations (3) and (6) we get for an intraclass 2×2 table as shown in Table 4 the unconditional probability

$$P(a,b,c) = \frac{n!}{a!\,b!\,c!}(p_{A\bar{A}})^a (p_{A\bar{A}})^b (p_{\bar{A}\bar{A}})^c \tag{10}$$

and the conditional probability, given the 'marginal' frequency r,

$$P(a|r) = \frac{n!\,r!\,s!\,2^b}{(2n)!\,a!\,b!\,c!}. \tag{11}$$

Now we see that the null hypothesis H_0 is equivalent to

$$H_0': \quad \Delta = 1,$$

where

$$\Delta = \frac{4p_{A\bar{A}}p_{\bar{A}\bar{A}}}{p_{A\bar{A}}^2} \tag{12}$$

gives a measure of dependence of two characteristics A and $\bar{A}$, in that $\Delta > 1$ corresponds to positive and $\Delta < 1$ to negative dependence.

In order to test the hypothesis H_0' we shall advocate the conditional test† based on the

† Tocher (1950) and Lehmann & Scheffé (1955) proved that Fisher's exact test, if appropriately randomized at the critical value, is the most powerful unbiased test of the hypothesis of independence with regard to the one-sided alternative. Our test, if randomized similarly, is also the most powerful unbiased test of the hypothesis (2), since the distribution (10) belongs to Lehmann & Scheffé's exponential family with $\theta = \log \Delta$ as the parameter tested and $\vartheta = \log (p_{A\bar{A}}/p_{\bar{A}\bar{A}})$ as a nuisance parameter. One of the authors, Goro Ishii, is now carrying on a study on the power function of this test, which will soon be published.

probability distribution (11) after the fashion of Fisher and Yates. That is, if we are concerned with positive dependence, $\Delta > 1$, as an alternative, we reject the null hypothesis whenever the frequency a satisfies $a \geqslant k$ for a suitably chosen constant k. For the critical value k we can calculate the size of the test by summing the probabilities (11) over the range $a \geqslant k$. On the other hand, if our concern lies in negative dependence, then the critical region becomes $a \leqslant k$. The use of the conditional distribution (11) was suggested by Armitage & Healy (1957)† in answer to a query concerning a practical example.

4. TABLES FOR THE CONDITIONAL TEST

We have constructed three tables useful in applying the conditional test as described in the preceding section to any numerical data of the intraclass 2×2 table. For the convenience of users we have followed the style of Finney (1948) and Latscha (1953) for the table of critical values and levels of significance and also that of Pearson (1947) for the order of accuracy of the normal approximation, both presented for an ordinary 2×2 contingency table.

We tabulate the critical values corresponding to the frequency a of the pair (A, A) and the levels of significance (or the sizes of the test) for either of two one-sided alternatives, positive and negative dependence, for the range $r = 4(1)s$, $s = 4(1)30$‡ of the marginal frequencies r and s and for the nominal levels of significance 0·10, 0·05, 0·01 and 0·001. Table 5 is available for the alternative of positive and Table 6 of negative dependence. Numbers in boldface show critical values and those in small type the corresponding exact levels of significance. When s is larger than 30 and our tables are of no use, a normal approximation is recommended. It is proved by Okamoto (1959) that if $n \to \infty$ and $r = 2p + o(\sqrt{n})$, $0 < p < 1$, then under the conditional distribution (11) the frequency a follows asymptotically a normal distribution, $N(2np^2, np^2q^2/4)$, where $q = 1-p$. Hence it holds that $4n^{\frac{3}{2}}(rs)^{-1}[a - r^2/(4n)]$ follows in the limit a standardized normal distribution. In order to get a better approximation it will be convenient to use Yates' correction

$$\Pr\{a \geqslant k|r\} \doteqdot \Pr\left\{Z > \frac{4n^{\frac{3}{2}}}{rs}\left(k - \frac{1}{2} - \frac{r^2}{4n}\right)\right\}, \tag{13}$$

where Z denotes a standardized normal variate. $\Pr\{a \leqslant k|r\}$ is obtained by reversing the two inequality signs and substituting $k + \frac{1}{2}$ for $k - \frac{1}{2}$ in (13). A better approximation is accomplished by standardizing a by its mean $\frac{1}{2}r(r-1)/(2n-1)$ and variance

$$\tfrac{1}{2}rs(r-1)(s-1)/\{(2n-1)^2(2n-3)\},$$

calculated from (7), that is,

$$\Pr\{a \geqslant k|r\} \doteqdot \Pr\left\{Z > \frac{(2n-1)\sqrt{[2(2n-s)]}}{\sqrt{[rs(r-1)(s-1)]}}\left[k - \frac{1}{2} - \frac{r(r-1)}{2(2n-1)}\right]\right\}. \tag{14}$$

The modification required for the reversed inequality is the same as with (13). The order of accuracy of the approximation (14) is illustrated by Table 7 for some typical values of the pair (r, s).

† The authors would like to thank the referee for drawing their attention to this work.

‡ The computation was actually carried out for the wider range $r = 4(1)s$, $s = 4(1)40$. Copies of this extension for $s > 30$ will be sent to interested readers on request.

5. An example

As an illustration let us examine the data† of the S blood groups (secretor S and non-secretor s) in saliva for 28 pairs of male dizygotic twins in Osaka City, Japan.

Pair table

	s (non-secretor)	S (secretor)	
s	5	9	$r = 19$
S	—	14	$s = 37$
			$n = 28$

The entry 5, for instance, denotes the frequency of pairs of twins of which both members belong to the s blood group. The marginal totals are

$$r = 5 \times 2 + 9 = 19 \quad \text{and} \quad s = 14 \times 2 + 9 = 37.$$

There appears a slight suspicion of positive dependence of S and s within a pair of twins.

Since $s = 37$ is larger than 30 and Table 5 is not available we use the normal approximation (14), obtaining

$$\Pr\{a \geqslant 5 | r = 19\} \doteqdot \Pr\{Z > 1{\cdot}167\} = 0{\cdot}1216.$$

It is seen from Table 7 (case $r = 20$, $s = 40$) that this figure cannot be far in error.‡ We therefore conclude that the data are insufficient for inferring the dependence in question.

With a sample of 23 pairs of twins, showing frequencies

	s	S	
s	6	5	$r = 17$
S	—	12	$s = 29$

it would have been possible to use Table 5. We find that for the conditional distribution, the probability of $a \geqslant 6$ is only 0·012 and we should now have fairly strong grounds for suspecting positive dependence of S and s within a pair of twins.

The authors are grateful to Prof. Junjiro Ogawa, Nihon University, and the referee for helpful suggestions.

References

Armitage, P. & Healy, M. J. R. (1957). Interpretation of χ^2 tests. *Biometrics*, **13**, 113–15.

Barnard, G. A. (1947). Significance tests for 2 × 2 tables. *Biometrika*, **34**, 123–38.

Finney, D. J. (1948). The Fisher–Yates test of significance in 2 × 2 contingency tables. *Biometrika*, **35**, 145–56.

Fisher, R. A. (1935). The logic of inductive inference. *J. R. Statist. Soc.* **98**, 39–54.

Fisher, R. A. (1955). Statistical methods and scientific induction. *J. R. Statist. Soc.* B, **17**, 69–78.

Haldane, J. B. S. (1939). The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. *Biometrika*, **31**, 346–55.

Latscha, R. (1953). Tests of significance in a 2 × 2 contingency table: extension of Finney's table. *Biometrika*, **40**, 74–86.

† They have been kindly lent to the authors by Prof. Mototsugu Kohama, Department of Anatomy, School of Medicine, Osaka University.

‡ The extended table mentioned in the footnote ‡ to p. 184, shows, in fact, that for independence the probability of having $a \geqslant 5$ with $r = 19$, $s = 37$ is more than 0·10.

LEHMANN, E. L. & SCHEFFÉ, H. (1955). Completeness, similar regions and unbiased estimation, II. *Sankhyā*, **15**, 219–36.

OKAMOTO, M. (1959). A convergence theorem for discrete probability distributions. *Ann. Inst. Statist. Math.* **11**, 107–12.

PEARSON, E. S. (1947). The choice of statistical test illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139–67.

PEARSON, E. S. (1955). Statistical concepts in their relation to reality. *J. R. Statist. Soc.* B, **17**, 204–7.

TOCHER, K. D. (1950). Extension of the Neyman–Pearson theory of tests to discontinuous variates. *Biometrika*, **37**, 130–44.

YATES, F. (1934). Contingency tables involving small numbers and the χ^2 test. *J. R. Statist. Soc.* Suppl. **1**, 217–35.

Table 5. *Significance levels of a for positive dependence*

(Values of a in bold type; corresponding probabilities P_a in small type.)

s	r	Probability 0·1	0·05	0·01	0·001
4	4	**2** ·086	—	—	—
6	6	**3** ·022	**3** ·022	—	—
	4	**2** ·048	**2** ·048	—	—
7	7	**3** ·082	—	—	—
8	8	**4** ·005+	**4** ·005+	**4** ·005+	—
	6	**3** ·012	**3** ·012	—	—
	4	**2** ·030	**2** ·030	—	—
9	9	**4** ·026	**4** ·026	—	—
	7	**3** ·049	**3** ·049	—	—
10	10	**4** ·070	**5** ·001+	**5** ·001+	—
	8	**4** ·003	**4** ·003	**4** ·003	—
	6	**3** ·007	**3** ·007	**3** ·007	—
	4	**2** ·021	**2** ·021	—	—
	2	**1** ·091	—	—	—
11	11	**5** ·008	**5** ·008	**5** ·008	—
	9	**4** ·015+	**4** ·015+	—	—
	7	**3** ·032	**3** ·032	—	—
	5	**2** ·077	—	—	—
12	12	**5** ·025−	**5** ·025−	**6** ·000	**6** ·0003
	10	**4** ·044	**4** ·044	**5** ·001−	**5** ·0007
	8	**3** ·082	**4** ·002	**4** ·002	—
	6	**3** ·005−	**3** ·005−	**3** ·005−	—
	4	**2** ·015+	**2** ·015+	—	—
	2	**1** ·077	—	—	—
13	13	**5** ·058	**6** ·002	**6** ·002	—
	11	**4** ·093	**5** ·004	**5** ·004	—
	9	**4** ·009	**4** ·009	**4** ·009	—
	7	**3** ·022	**3** ·022	—	—
	5	**2** ·059	—	—	—
14	14	**6** ·008	**6** ·008	**6** ·008	**7** ·0001−
	12	**5** ·015+	**5** ·015+	**6** ·000	**6** ·0002
	10	**4** ·029	**4** ·029	**5** ·000	**5** ·0004
	8	**3** ·059	**4** ·001+	**4** ·001+	—
	6	**3** ·003	**3** ·003	**3** ·003	—
	4	**2** ·012	**2** ·012	—	—
	2	**1** ·067	—	—	—
15	15	**6** ·022	**6** ·022	**7** ·001−	**7** ·0007
	13	**5** ·037	**5** ·037	**6** ·001+	—
	11	**4** ·065−	**5** ·003	**5** ·003	—
	9	**4** ·006	**4** ·006	**4** ·006	—
	7	**3** ·015+	**3** ·015+	—	—
	5	**2** ·046	**2** ·046	—	—
16	16	**6** ·048	**6** ·048	**7** ·003	**8** ·0000
	14	**5** ·074	**6** ·005+	**6** ·005+	**7** ·0000
	12	**5** ·010−	**5** ·010−	**5** ·010−	**6** ·0001−
	10	**4** ·020	**4** ·020	**5** ·000	**5** ·0002
	8	**3** ·044	**3** ·044	**4** ·001−	**4** ·0007
16	6	**3** ·002	**3** ·002	**3** ·002	—
	4	**2** ·009	**2** ·009	**2** ·009	—
	2	**1** ·059	—	—	—
17	17	**6** ·087	**7** ·008	**7** ·008	**8** ·0002
	15	**6** ·014	**6** ·014	**7** ·000	**7** ·0004
	13	**5** ·025−	**5** ·025−	**6** ·001−	**6** ·0008
	11	**4** ·046	**4** ·046	**5** ·002	—
	9	**3** ·092	**4** ·004	**4** ·004	—
	7	**3** ·011	**3** ·011	—	—
	5	**2** ·036	**2** ·036	—	—
18	18	**7** ·019	**7** ·019	**8** ·001−	**8** ·0009
	16	**6** ·031	**6** ·031	**7** ·002	**8** ·0000
	14	**5** ·052	**6** ·003	**6** ·003	**7** ·0000
	12	**4** ·090	**5** ·006	**5** ·006	**6** ·0001−
	10	**4** ·014	**4** ·014	**5** ·000	**5** ·0002
	8	**3** ·033	**3** ·033	**4** ·000	**4** ·0005−
	6	**2** ·090	**3** ·002	**3** ·002	—
	4	**2** ·008	**2** ·008	**2** ·008	—
	2	**1** ·053	—	—	—
19	19	**7** ·039	**7** ·039	**8** ·003	**9** ·0001−
	17	**6** ·060	**7** ·005−	**7** ·005−	**8** ·0001+
	15	**5** ·094	**6** ·009	**6** ·009	**7** ·0002
	13	**5** ·017	**5** ·017	**6** ·000	**6** ·0005−
	11	**4** ·034	**4** ·034	**5** ·001+	—
	9	**3** ·072	**4** ·003	**4** ·003	—
	7	**3** ·009	**3** ·009	**3** ·009	—
	5	**2** ·031	**2** ·031	—	—
20	20	**7** ·069	**8** ·008	**8** ·008	**9** ·0003
	18	**7** ·012	**7** ·012	**8** ·000	**8** ·0005−
	16	**6** ·021	**6** ·021	**7** ·001−	**7** ·0010−
	14	**5** ·037	**5** ·037	**6** ·002	**7** ·0000
	12	**4** ·068	**5** ·004	**5** ·004	**6** ·0000
	10	**4** ·010+	**4** ·010+	**5** ·000	**5** ·0001−
	8	**3** ·026	**3** ·026	**4** ·000	**4** ·0003
	6	**2** ·076	**3** ·001+	**3** ·001+	—
	4	**2** ·006	**2** ·006	**2** ·006	—
	2	**1** ·048	**1** ·048	—	—
21	21	**8** ·017	**8** ·017	**9** ·001−	**9** ·0010−
	19	**7** ·026	**7** ·026	**8** ·002	**9** ·0000
	17	**6** ·042	**6** ·042	**7** ·003	**8** ·0001−
	15	**5** ·069	**6** ·006	**6** ·006	**7** ·0001+
	13	**5** ·012	**5** ·012	**6** ·000	**6** ·0003
	11	**4** ·026	**4** ·026	**5** ·001−	**5** ·0007
	9	**3** ·058	**4** ·002	**4** ·002	—
	7	**3** ·007	**3** ·007	**3** ·007	—
	5	**2** ·026	**2** ·026	—	—
22	22	**8** ·032	**8** ·032	**9** ·003	**10** ·0001−
	20	**7** ·048	**7** ·048	**8** ·005−	**9** ·0002
	18	**6** ·074	**7** ·008	**7** ·008	**8** ·0003
	16	**6** ·015−	**6** ·015−	**7** ·001−	**7** ·0006
	14	**5** ·027	**5** ·027	**6** ·001+	**7** ·0000
	12	**4** ·053	**5** ·003	**5** ·003	**6** ·0000

Table 5 (*cont.*)

s	r	Probability				s	r	Probability			
		0·1	0·05	0·01	0·001			0·1	0·05	0·01	0·001
22	10	4 ·008	4 ·008	4 ·008	5 ·0001⁻	27	25	8 ·099	9 ·018	10 ·002	11 ·0001⁻
	8	3 ·021	3 ·021	4 ·000	4 ·0002		23	8 ·028	8 ·028	9 ·003	10 ·0001⁺
	6	2 ·065⁻	3 ·001⁻	3 ·001⁻	3 ·0010⁻		21	7 ·043	7 ·043	8 ·005⁻	9 ·0002
	4	2 ·005⁺	2 ·005⁺	2 ·005⁺	—		19	6 ·069	7 ·009	7 ·009	8 ·0004
	2	1 ·043	1 ·043	—	—		17	6 ·016	6 ·016	7 ·001⁻	7 ·0009
							15	5 ·031	5 ·031	6 ·002	7 ·0000
23	23	8 ·056	9 ·007	9 ·007	10 ·0003		13	4 ·061	5 ·005⁻	5 ·005⁻	6 ·0001⁻
	21	7 ·081	8 ·011	9 ·001⁻	9 ·0006		11	4 ·012	4 ·012	5 ·000	5 ·0003
	19	7 ·018	7 ·018	8 ·001⁺	9 ·0000		9	3 ·032	3 ·032	4 ·001⁻	4 ·0009
	17	6 ·030	6 ·030	7 ·002	8 ·0000		7	2 ·096	3 ·004	3 ·004	—
	15	5 ·052	6 ·004	6 ·004	7 ·0001⁻		5	2 ·017	2 ·017	—	—
	13	4 ·094	5 ·009	5 ·009	6 ·0002						
	11	4 ·020	4 ·020	5 ·001⁻	5 ·0005⁺	28	28	10 ·022	10 ·022	11 ·002	12 ·0001⁺
	9	3 ·047	3 ·047	4 ·002	—		26	9 ·032	9 ·032	10 ·004	11 ·0002
	7	3 ·005⁺	3 ·005⁺	3 ·005⁺	—		24	8 ·047	8 ·047	9 ·006	10 ·0004
	5	2 ·022	2 ·022	—	—		22	7 ·071	8 ·010⁺	9 ·001⁻	9 ·0007
							20	7 ·018	7 ·018	8 ·001⁺	9 ·0000
24	24	8 ·089	9 ·014	10 ·001⁻	10 ·0010⁻		18	6 ·031	6 ·031	7 ·003	8 ·0001⁻
	22	8 ·022	8 ·022	9 ·002	10 ·0000		16	5 ·055⁺	6 ·005⁺	6 ·005⁺	7 ·0002
	20	7 ·034	7 ·034	8 ·003	9 ·0001⁻		14	5 ·012	5 ·012	6 ·000	6 ·0004
	18	6 ·055⁻	7 ·005⁺	7 ·005⁺	8 ·0002		12	4 ·026	4 ·026	5 ·001⁺	6 ·0000
	16	5 ·089	6 ·010⁺	7 ·000	7 ·0004		10	3 ·063	4 ·003	4 ·003	5 ·0000
	14	5 ·020	5 ·020	6 ·001⁻	6 ·0009		8	3 ·011	3 ·011	4 ·000	4 ·0001⁺
	12	4 ·041	4 ·041	5 ·002	6 ·0000		6	2 ·043	2 ·043	3 ·001⁻	3 ·0005⁺
	10	3 ·089	4 ·006	4 ·006	5 ·0000		4	2 ·003	2 ·003	2 ·003	—
	8	3 ·017	3 ·017	4 ·000	4 ·0002		2	1 ·034	1 ·034	—	—
	6	2 ·056	3 ·001⁻	3 ·001⁻	3 ·0008						
	4	2 ·004	2 ·004	2 ·004	—	29	29	10 ·037	10 ·037	11 ·005⁺	12 ·0004
	2	1 ·040	1 ·040	—	—		27	9 ·052	10 ·008	10 ·008	11 ·0006
							25	8 ·075⁻	9 ·012	10 ·001⁺	11 ·0000
25	25	9 ·026	9 ·026	10 ·003	11 ·0001⁺		23	8 ·020	8 ·020	9 ·002	10 ·0001⁻
	23	8 ·039	8 ·039	9 ·004	10 ·0002		21	7 ·032	7 ·032	8 ·003	9 ·0001⁺
	21	7 ·059	8 ·007	8 ·007	9 ·0003		19	6 ·054	7 ·006	7 ·006	8 ·0003
	19	6 ·090	7 ·012	8 ·001⁻	8 ·0007		17	5 ·090	6 ·012	7 ·001⁻	7 ·0006
	17	6 ·022	6 ·022	7 ·001⁺	8 ·0000		15	5 ·024	5 ·024	6 ·001⁺	7 ·0000
	15	5 ·040	5 ·040	6 ·003	7 ·0001⁻		13	4 ·050⁺	5 ·004	5 ·004	6 ·0001⁻
	13	4 ·075⁺	5 ·006	5 ·006	6 ·0001⁺		11	4 ·010⁻	4 ·010⁻	4 ·010⁻	5 ·0002
	11	4 ·015⁺	4 ·015⁺	5 ·000	5 ·0004		9	3 ·027	3 ·027	4 ·001⁻	4 ·0007
	9	3 ·039	3 ·039	4 ·001⁺	—		7	2 ·085⁻	3 ·003	3 ·003	—
	7	3 ·004	3 ·004	3 ·004	—		5	2 ·015⁻	2 ·015⁻	—	—
	5	2 ·019	2 ·019	—	—		3	1 ·097	—	—	—
26	26	9 ·045⁺	9 ·045⁺	10 ·006	11 ·0003	30	30	10 ·058	11 ·010⁻	11 ·010⁻	12 ·0009
	24	8 ·065⁻	9 ·009	9 ·009	10 ·0006		28	9 ·080	10 ·015⁻	11 ·001⁺	12 ·0001⁻
	22	7 ·093	8 ·015⁻	9 ·001⁺	10 ·0000		26	9 ·023	9 ·023	10 ·002	11 ·0001⁺
	20	7 ·024	7 ·024	8 ·002	9 ·0001⁻		24	8 ·035⁻	8 ·035⁻	9 ·004	10 ·0002
	18	6 ·041	6 ·041	7 ·004	8 ·0001⁺		22	7 ·054	8 ·007	8 ·007	9 ·0004
	16	5 ·063	6 ·007	6 ·007	7 ·0003		20	6 ·086	7 ·013	8 ·001⁻	8 ·0009
	14	5 ·015⁺	5 ·015⁺	6 ·001⁻	6 ·0006		18	6 ·024	6 ·024	7 ·002	8 ·0000
	12	4 ·033	4 ·033	5 ·002	6 ·0000		16	5 ·044	5 ·044	6 ·004	7 ·0001⁺
	10	3 ·075⁻	4 ·004	4 ·004	5 ·0000		14	4 ·085⁻	5 ·009	5 ·009	6 ·0003
	8	3 ·014	3 ·014	4 ·000	4 ·0001⁺		12	4 ·021	4 ·021	5 ·001⁻	5 ·0009
	6	2 ·049	2 ·049	3 ·001⁻	3 ·0006		10	3 ·054	4 ·003	4 ·003	5 ·0000
	4	2 ·004	2 ·004	2 ·004	—		8	3 ·010⁻	3 ·010⁻	3 ·010⁻	4 ·0001⁻
	2	1 ·037	1 ·037	—	—		6	2 ·038	2 ·038	3 ·000	3 ·0004
							4	2 ·003	2 ·003	2 ·003	—
27	27	9 ·072	10 ·012	11 ·001⁻	11 ·0010⁻		2	1 ·032	1 ·032	—	—

Table 6. *Significance levels of a for negative dependence*

(Values of a in bold type; corresponding probabilities P_a in small type.)

s	r	Probability 0·1	0·05	0·01	0·001
6	6	**0** ·069	—	—	—
7	7	**0** ·037	**0** ·037	—	—
8	8	**0** ·020	**0** ·020	—	—
9	9	**0** ·011	**0** ·011	—	—
	7	**0** ·090	—	—	—
10	10	**0** ·006	**0** ·006	**0** ·006	—
	8	**0** ·053	—	—	—
11	11	**1** ·083	**0** ·003	**0** ·003	—
	9	**0** ·030	**0** ·030	—	—
12	12	**1** ·052	**0** ·002	**0** ·002	—
	10	**0** ·017	**0** ·017	—	—
	8	**0** ·091	—	—	—
13	13	**1** ·032	**1** ·032	**0** ·001⁻	**0** ·0008
	11	**0** ·010⁻	**0** ·010⁻	**0** ·010⁻	—
	9	**0** ·057	—	—	—
14	14	**1** ·019	**1** ·019	**0** ·000	**0** ·0004
	12	**1** ·096	**0** ·006	**0** ·006	—
	10	**0** ·034	**0** ·034	—	—
15	15	**1** ·011	**1** ·011	**0** ·000	**0** ·0002
	13	**1** ·063	**0** ·003	**0** ·003	—
	11	**0** ·021	**0** ·021	—	—
	9	**0** ·086	—	—	—
16	16	**2** ·081	**1** ·007	**1** ·007	**0** ·0001⁺
	14	**1** ·040	**1** ·040	**0** ·002	—
	12	**0** ·012	**0** ·012	—	—
	10	**0** ·055⁺	—	—	—
17	17	**2** ·054	**1** ·004	**1** ·004	**0** ·0001⁻
	15	**1** ·025⁺	**1** ·025⁺	**0** ·001⁻	**0** ·0009
	13	**0** ·007	**0** ·007	**0** ·007	—
	11	**0** ·035⁻	**0** ·035⁻	—	—
18	18	**2** ·035⁺	**2** ·035⁺	**1** ·002	**0** ·0000
	16	**1** ·016	**1** ·016	**0** ·001⁻	**0** ·0005⁺
	14	**1** ·067	**0** ·004	**0** ·004	—
	12	**0** ·022	**0** ·022	—	—
	10	**0** ·078	—	—	—
19	19	**2** ·023	**2** ·023	**1** ·001⁺	**0** ·0000
	17	**2** ·091	**1** ·010⁻	**1** ·010⁻	**0** ·0003
	15	**1** ·044	**1** ·044	**0** ·002	—
	13	**0** ·013	**0** ·013	—	—
	11	**0** ·051	—	—	—
20	20	**2** ·015⁻	**2** ·015⁻	**1** ·001⁻	**1** ·0007
	18	**2** ·063	**1** ·006	**1** ·006	**0** ·0001⁺
	16	**1** ·029	**1** ·029	**0** ·001⁺	—
	14	**1** ·099	**0** ·008	**0** ·008	—
	12	**0** ·033	**0** ·033	—	—
21	21	**3** ·075⁺	**2** ·009	**2** ·009	**1** ·0004
	19	**2** ·042	**2** ·042	**1** ·003	**0** ·0001⁻
	17	**1** ·018	**1** ·018	**0** ·001⁻	**0** ·0008
	15	**1** ·068	**0** ·005⁻	**0** ·005⁻	—
	13	**0** ·021	**0** ·021	—	—
	11	**0** ·069	—	—	—
22	22	**3** ·052	**2** ·006	**2** ·006	**1** ·0002
	20	**2** ·028	**2** ·028	**1** ·002	**0** ·0000
	18	**2** ·096	**1** ·012	**0** ·000	**0** ·0004
	16	**1** ·046	**1** ·046	**0** ·003	—
	14	**0** ·013	**0** ·013	—	—
	12	**0** ·046	**0** ·046	—	—
23	23	**3** ·036	**3** ·036	**2** ·004	**1** ·0001⁺
	21	**2** ·018	**2** ·018	**1** ·001⁺	**0** ·0000
	19	**2** ·067	**1** ·007	**1** ·007	**0** ·0002
	17	**1** ·030	**1** ·030	**0** ·002	—
	15	**1** ·094	**0** ·008	**0** ·008	—
	13	**0** ·030	**0** ·030	—	—
	11	**0** ·089	—	—	—
24	24	**3** ·024	**3** ·024	**2** ·002	**1** ·0001⁻
	22	**3** ·083	**2** ·012	**1** ·001⁻	**1** ·0007
	20	**2** ·046	**2** ·046	**1** ·004	**0** ·0001⁺
	18	**1** ·020	**1** ·020	**0** ·001⁻	**0** ·0010⁻
	16	**1** ·066	**0** ·005⁺	**0** ·005⁺	—
	14	**0** ·020	**0** ·020	—	—
	12	**0** ·061	—	—	—
25	25	**4** ·094	**3** ·016	**2** ·001⁺	**1** ·0000
	23	**3** ·059	**2** ·008	**2** ·008	**1** ·0004
	21	**2** ·031	**2** ·031	**1** ·003	**0** ·0001⁻
	19	**2** ·096	**1** ·013	**0** ·001⁻	**0** ·0006
	17	**1** ·045⁻	**1** ·045⁻	**0** ·003	—
	15	**0** ·013	**0** ·013	—	—
	13	**0** ·041	**0** ·041	—	—
26	26	**4** ·068	**3** ·011	**2** ·001⁻	**2** ·0008
	24	**3** ·041	**3** ·041	**2** ·005⁻	**1** ·0002
	22	**2** ·021	**2** ·021	**1** ·002	**0** ·0000
	20	**2** ·068	**1** ·008	**1** ·008	**0** ·0003
	18	**1** ·030	**1** ·030	**0** ·002	—
	16	**1** ·088	**0** ·008	**0** ·008	—
	14	**0** ·027	**0** ·027	—	—
	12	**0** ·076	—	—	—

Table 6 (*cont.*)

s	r	Probability 0·1	0·05	0·01	0·001	s	r	Probability 0·1	0·05	0·01	0·001
27	27	4 ·049	4 ·049	3 ·007	2 ·0005⁻	29	29	4 ·024	4 ·024	3 ·003	2 ·0002
	25	3 ·028	3 ·028	2 ·003	1 ·0001⁺		27	4 ·075⁻	3 ·013	2 ·001⁺	1 ·0000
	23	3 ·087	2 ·014	1 ·001⁺	0 ·0000		25	3 ·045⁻	3 ·045⁻	2 ·006	1 ·0004
	21	2 ·048	2 ·048	1 ·005⁺	0 ·0002		23	2 ·023	2 ·023	1 ·002	0 ·0001⁻
	19	1 ·020	1 ·020	0 ·001⁺	—		21	2 ·068	1 ·009	1 ·009	0 ·0004
	17	1 ·062	0 ·005⁺	0 ·005⁺	—		19	1 ·029	1 ·029	0 ·002	—
	15	0 ·018	0 ·018	—	—		17	1 ·081	0 ·008	0 ·008	—
	13	0 ·053	—	—	—		15	0 ·024	0 ·024	—	—
							13	0 ·065⁺	—	—	—
28	28	4 ·034	4 ·034	3 ·004	2 ·0003	30	30	5 ·083	4 ·016	3 ·002	2 ·0001⁻
	26	3 ·019	3 ·019	2 ·002	1 ·0001⁻		28	4 ·054	3 ·009	3 ·009	2 ·0007
	24	3 ·063	2 ·009	2 ·009	1 ·0006		26	3 ·031	3 ·031	2 ·004	1 ·0002
	22	2 ·033	2 ·033	1 ·003	0 ·0001⁺		24	3 ·089	2 ·015⁺	1 ·001⁺	0 ·0000
	20	2 ·094	1 ·013	0 ·001⁻	0 ·0007		22	2 ·048	2 ·048	1 ·006	0 ·0002
	18	1 ·043	1 ·043	0 ·003	—		20	1 ·020	1 ·020	0 ·001⁺	—
	16	0 ·012	0 ·012	—	—		18	1 ·058	0 ·005⁻	0 ·005⁻	—
	14	0 ·036	0 ·036	—	—		16	0 ·016	0 ·016	—	—
	12	0 ·092	—	—	—		14	0 ·046	0 ·046	—	—

Table 7. *Normal approximation for significance levels*

Chance that	True	Normal approx.	Chance that	True	Normal approx.	Chance that	True	Normal approx.	Chance that	True	Normal approx.
$r = 20, s = 40$			$r = 30, s = 40$			$r = 30, s = 60$			$r = 30, s = 100$		
$a \leqslant$ 0	0·0075	0·0130	$a \leqslant$ 2	0·0033	0·0045	$a \leqslant$ 0	0·0006	0·0017	$a \leqslant$ 0	0·0124	0·0217
1	·0724	·0796	3	·0242	·0273	1	·0080	·0117	1	·0871	·0934
2	·2793	·2778	4	·1042	·1080	2	·0496	·0551	2	·2779	·2706
			5	·2890	·2907	3	·1747	·1766			
$a \geqslant$ 5	0·1438	0·1476	$a \geqslant$ 8	0·2028	0·2062	$a \geqslant$ 7	0·1380	0·1404	$a \geqslant$ 5	0·2094	0·2147
6	·0300	·0311	9	·0631	·0661	8	·0398	·0403	6	·0704	·0680
7	·0033	·0036	10	·0126	·0142	9	·0077	·0078	7	·0167	·0142
			11	·0015	·0020	10	·0010	·0010	8	·0027	·0019
$r = 40, s = 40$			$r = 40, s = 60$			$r = 40, s = 100$			$r = 80, s = 100$		
$a \leqslant$ 5	0·0024	0·0030	$a \leqslant$ 3	0·0039	0·0051	$a \leqslant$ 0	0·0003	0·0014	$a \leqslant$ 10	0·0010	0·0012
6	·0153	·0170	4	·0212	·0237	1	·0047	·0080	11	·0039	·0044
7	·0650	·0679	5	·0782	·0814	2	·0285	·0342	12	·0132	·0141
8	·1912	·1940	6	·2075	·2092	3	·1041	·1082	13	·0372	·0385
						4	·2602	·2576	14	·0882	·0897
$a \geqslant$ 12	0·1503	0·1533	$a \geqslant$ 10	0·1683	0·1707				15	·1783	·1796
13	·0467	·0494	11	·0603	·0620	$a \geqslant$ 8	0·1333	0·1344			
14	·0099	·0113	12	·0158	·0168	9	·0464	·0454	$a \geqslant$ 20	0·2146	0·2159
15	·0014	·0018	13	·0030	·0033	10	·0123	·0114	21	·1115	·1128
						11	·0024	·0021	22	·0497	·0507
									23	·0188	·0195
									24	·0060	·0064
									25	·0016	·0018

Biometrika (1961), **48**, 1 *and* 2, *p.* 191
Printed in Great Britain

Miscellanea

Table of Neyman-shortest unbiased confidence intervals for the Poisson parameter†

BY COLIN R. BLYTH AND DAVID W. HUTCHINSON
University of Illinois, Urbana, Illinois

1. *Introduction and summary.* Neyman-shortest unbiased confidence intervals for the Poisson parameter can be obtained in exactly the same way as was done for the Binomial parameter by Blyth & Hutchinson (1960). In the Poisson case the equations for the best unbiased acceptance region of size α, namely $n_0+\gamma_0 \leqslant X+Y \leqslant n_1+\gamma_1$ reduce to

$$n_0, n_1 \text{ integers}, \quad 0 \leqslant \gamma_0 \leqslant 1, \quad 0 \leqslant \gamma_1 \leqslant 1,$$

$$\gamma_0 = \frac{(n_1-\lambda^*)\{P_{\lambda^*}(n_0 \leqslant X \leqslant n_1-1)-\alpha\}-n_0 P_{\lambda^*}(X=n_0)+n_1 P_{\lambda^*}(X=n_1)}{(n_1-n_0)P_{\lambda^*}(X=n_0)},$$

$$\gamma_1 = \frac{(n_0-\lambda^*)\{P_{\lambda^*}(n_0 \leqslant X \leqslant n_1-1)-\alpha\}-n_0 P_{\lambda^*}(X=n_0)+n_1 P_{\lambda^*}(X=n_1)}{(n_1-n_0)P_{\lambda^*}(X=n_1)}.$$

The University of Illinois Digital Computer Laboratory's ILLIAC was programmed to compute and print out $n_0+\gamma_0$, $n_1+\gamma_1$ for any α and any set of λ^* values. The values used were $\alpha = 0{\cdot}95$, $0{\cdot}99$ and $\lambda^* = 0{\cdot}005$, $0{\cdot}015$, ..., $0{\cdot}535$, $0{\cdot}55$, $0{\cdot}65$, ..., $74{\cdot}95$, $75{\cdot}5$, $76{\cdot}5$, ..., $300{\cdot}5$. Neyman-shortest unbiased confidence sets were then read off from this print-out.

2. *Use of the tables.* Let X be a Poisson variable having expectation λ and Y be Rectangular (0, 1) variable, with X and Y independent. If $X = x$ and $Y = y$ (using a table of random numbers) are observed, enter the table for α at the location $x+y$ and read there the observed α-confidence interval for λ. Tables are given for $\alpha = 0{\cdot}95$, $0{\cdot}99$ and for values of $X+Y$ as follows:

$$0{\cdot}01\,(0{\cdot}01)\,0{\cdot}10\,(0{\cdot}02)\,0{\cdot}20\,(0{\cdot}05)\,1{\cdot}00\,(0{\cdot}1)\,10{\cdot}0\,(0{\cdot}2)\,40{\cdot}0\,(0{\cdot}5)\,55{\cdot}0\,(1)\,250.$$

The intervals at which $X+Y$ is given and the number of places given in the confidence limits are so chosen that interpolation and round-off errors are small compared to the length of the interval (less than 2% except for a few $X+Y$ values below 1). For large values of X the Normal approximation to the Poisson distribution gives the approximate α-confidence interval $\frac{1}{2}\{c^2+2X \pm c(c^2+4X)^{\frac{1}{2}}\}$, where c is the $1-\frac{1}{2}\alpha$ point of the standardized Normal (0, 1) distribution. For $X > 250$ the differences between these limits and the Neyman-shortest unbiased limits are small compared to the length of the interval (less than 1%).

Example 1. If X is Poisson (expectation λ) and $X = 23$, $Y = 0{\cdot}65$ are observed, $0{\cdot}99$ confidence limits for λ are $12{\cdot}7$, $37{\cdot}9$. For $X = 23$, depending on the observed Y, the lower limit will fall between $12{\cdot}4$ and $13{\cdot}1$ and the upper limit between $37{\cdot}0$ and $38{\cdot}3$, with an average length of $25{\cdot}0$. The usual equal probability tail interval first given by Garwood (1936) and tabled in *Biometrika Tables for Statisticians*, **1**, Table 40 is $12{\cdot}5$, $38{\cdot}5$ for $X = 23$, with length $26{\cdot}0$.

Example 2. If $X_1, \ldots, X_{20}$ are 20 independent Poisson variables having a common expectation λ, then ΣX_i is Poisson (expectation 20λ). If $\Sigma X_i = 78$ and $Y = 0{\cdot}54$ are observed, the observed $0{\cdot}95$ confidence interval for 20λ is 62, 96; the observed $0{\cdot}95$ confidence interval for λ is therefore $3{\cdot}1$, $4{\cdot}8$.

REFERENCES

BLYTH, C. R. & HUTCHINSON, D. W. (1960). Table of Neyman-shortest unbiased confidence intervals for the binomial parameter. *Biometrika*, **47**, 381–91.
GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, **28**, 437–42.

† Work supported by the National Science Foundation.

Table 1. *Neyman-shortest unbiased 0·95 and 0·99 confidence intervals for λ based on X, Poisson (λ) and an independent Y, Rectangular (0, 1)*

$X+Y$	$\alpha=0{\cdot}95$		$\alpha=0{\cdot}99$		$X+Y$	$\alpha=0{\cdot}95$		$\alpha=0{\cdot}99$		$X+Y$	$\alpha=0{\cdot}95$		$\alpha=0{\cdot}99$	
0·01	0	0	0	0	3·6	0·5	8·1	0·2	10·3	9·1	4·0	15·8	2·9	18·6
·02	0	0	0	·9	3·7	·6	8·2	·3	10·5	9·2	4·0	15·9	3·0	18·8
·03	0	0	0	1·3	3·8	·6	8·4	·3	10·6	9·3	4·1	16·1	3·0	18·9
·04	0	0	0	1·7	3·9	·7	8·5	·4	10·7	9·4	4·1	16·2	3·1	19·1
·05	0	0	0	1·9	4·0	·8	8·6	·4	10·8	9·5	4·2	16·3	3·1	19·2
0·06	0	0·3	0	2·1	4·1	0·8	8·8	0·5	11·0	9·6	4·3	16·5	3·2	19·4
·07	0	·5	0	2·3	4·2	·9	9·0	·5	11·3	9·7	4·4	16·6	3·3	19·5
·08	0	·7	0	2·4	4·3	·9	9·1	·5	11·5	9·8	4·4	16·7	3·3	19·6
·09	0	·8	0	2·5	4·4	1·0	9·3	·5	11·6	9·9	4·5	16·8	3·4	19·7
·10	0	·9	0	2·7	4·5	1·0	9·5	·5	11·8	10·0	4·6	16·9	3·5	19·8
0·12	0	1·2	0	2·9	4·6	1·1	9·6	0·6	11·9	10·2	4·7	17·2	3·6	20·2
·14	0	1·3	0	3·0	4·7	1·1	9·7	·6	12·1	10·4	4·8	17·5	3·7	20·5
·16	0	1·5	0	3·2	4·8	1·2	9·8	·7	12·2	10·6	5·0	17·8	3·8	20·8
·18	0	1·6	0	3·3	4·9	1·3	10·0	·7	12·3	10·8	5·1	18·0	4·0	21·0
·20	0	1·7	0	3·4	5·0	1·4	10·1	·8	12·4	11·0	5·3	18·2	4·1	21·2
0·25	0	2·0	0	3·6	5·1	1·4	10·3	0·8	12·6	11·2	5·4	18·5	4·2	21·6
·30	0	2·2	0	3·8	5·2	1·4	10·4	·9	12·9	11·4	5·6	18·8	4·3	21·9
·35	0	2·4	0	4·0	5·3	1·5	10·6	·9	13·0	11·6	5·7	19·1	4·4	22·1
·40	0	2·5	0	4·1	5·4	1·5	10·8	·9	13·2	11·8	5·8	19·3	4·6	22·4
·45	0	2·6	0	4·3	5·5	1·6	10·9	1·0	13·4	12·0	6·0	19·5	4·8	22·6
0·50	0	2·8	0	4·4	5·6	1·6	11·0	1·0	13·5	12·2	6·1	19·8	4·9	22·9
·55	0	2·9	0	4·5	5·7	1·7	11·2	1·1	13·6	12·4	6·3	20·1	5·0	23·3
·60	0	3·0	0	4·6	5·8	1·8	11·3	1·1	13·8	12·6	6·4	20·3	5·1	23·5
·65	0	3·0	0	4·7	5·9	1·9	11·4	1·2	13·9	12·8	6·6	20·6	5·2	23·8
·70	0	3·1	0	4·7	6·0	2·0	11·5	1·3	14·0	13·0	6·8	20·8	5·4	24·0
0·75	0	3·2	0	4·8	6·1	2·0	11·7	1·3	14·2	13·2	6·9	21·1	5·5	24·3
·80	0	3·3	0	4·9	6·2	2·1	11·8	1·3	14·4	13·4	7·0	21·4	5·6	24·6
·85	0	3·3	0	4·9	6·3	2·1	12·0	1·4	14·6	13·6	7·1	21·6	5·7	24·9
·90	0	3·4	0	5·0	6·4	2·2	12·2	1·4	14·7	13·8	7·3	21·8	5·9	25·1
1·0	0	3·5	0	5·1	6·5	2·2	12·3	1·4	14·9	14·0	7·5	22·1	6·0	25·3
1·1	0	3·8	0	5·5	6·6	2·3	12·4	1·5	15·0	14·2	7·6	22·4	6·1	25·7
1·2	0	4·1	0	5·9	6·7	2·4	12·6	1·6	15·1	14·4	7·7	22·6	6·3	26·0
1·3	0	4·3	0	6·1	6·8	2·4	12·7	1·6	15·3	14·6	7·9	22·9	6·4	26·2
1·4	0	4·5	0	6·4	6·9	2·5	12·8	1·7	15·4	14·8	8·0	23·1	6·5	26·4
1·5	0	4·7	0	6·6	7·0	2·6	12·9	1·8	15·5	15·0	8·2	23·3	6·7	26·7
1·6	0	4·9	0	6·7	7·1	2·7	13·1	1·8	15·7	15·2	8·3	23·6	6·8	27·0
1·7	0	5·0	0	6·9	7·2	2·7	13·2	1·9	15·9	15·4	8·5	23·9	6·9	27·3
1·8	0	5·2	0	7·0	7·3	2·8	13·4	1·9	16·0	15·6	8·6	24·1	7·1	27·5
1·9	0	5·3	0	7·1	7·4	2·8	13·5	1·9	16·2	15·8	8·8	24·4	7·2	27·8
2·0	0	5·4	0	7·2	7·5	2·9	13·7	2·0	16·4	16·0	9·0	24·6	7·4	28·0
2·1	0·1	5·6	0	7·6	7·6	2·9	13·8	2·0	16·5	16·2	9·1	24·9	7·5	28·3
2·2	·1	5·9	0	7·8	7·7	3·0	13·9	2·1	16·6	16·4	9·2	25·1	7·6	28·6
2·3	·1	6·1	0	8·1	7·8	3·1	14·0	2·2	16·7	16·6	9·4	25·4	7·7	28·9
2·4	·1	6·2	0	8·3	7·9	3·2	14·1	2·2	16·8	16·8	9·5	25·6	7·9	29·1
2·5	·1	6·4	0	8·4	8·0	3·3	14·3	2·3	17·0	17·0	9·7	25·8	8·1	29·3
2·6	0·1	6·6	0	8·6	8·1	3·3	14·4	2·4	17·2	17·2	9·9	26·1	8·2	29·6
2·7	·2	6·7	0	8·7	8·2	3·4	14·6	2·4	17·3	17·4	10·0	26·4	8·3	29·9
2·8	·2	6·8	0	8·8	8·3	3·4	14·7	2·4	17·5	17·6	10·2	26·6	8·4	30·3
2·9	·3	6·9	0·1	9·0	8·4	3·5	14·9	2·5	17·7	17·8	10·3	26·8	8·6	30·4
3·0	·4	7·1	·1	9·1	8·5	3·6	15·0	2·5	17·8	18·0	10·5	27·1	8·8	30·6
3·1	0·4	7·3	0·2	9·4	8·6	3·6	15·1	2·6	17·9	18·2	10·6	27·3	8·9	30·9
3·2	·4	7·5	·2	9·6	8·7	3·7	15·3	2·7	18·1	18·4	10·8	27·6	9·0	31·2
3·3	·4	7·6	·2	9·8	8·8	3·8	15·4	2·7	18·2	18·6	10·9	27·8	9·1	31·5
3·4	·5	7·8	·2	10·0	8·9	3·8	15·5	2·8	18·3	18·8	11·1	28·1	9·3	31·7
3·5	·5	8·0	·2	10·2	9·0	3·9	15·6	2·9	18·4	19·0	11·3	28·3	9·5	31·9

Note: the pairs of figures under each α-heading are lower and upper confidence limits for λ.

$X+Y$	$\alpha=0{\cdot}95$		$\alpha=0{\cdot}99$		$X+Y$	$\alpha=0{\cdot}95$		$\alpha=0{\cdot}99$		$X+Y$	$\alpha=0{\cdot}95$		$\alpha=0{\cdot}99$	
19·2	11·4	28·6	9·6	32·2	30·2	20·2	41·7	17·7	46·1	43·0	31·0	56·6	27·8	61·6
19·4	11·5	28·8	9·7	32·5	30·4	20·4	42·0	17·9	46·4	43·5	31·3	57·2	28·2	62·2
19·6	11·7	29·1	9·8	32·8	30·6	20·5	42·2	18·0	46·6	44·0	31·8	57·8	28·6	62·8
19·8	11·9	29·3	10·0	33·0	30·8	20·7	42·4	18·2	46·8	44·5	32·2	58·4	29·0	63·4
20·0	12·0	29·5	10·2	33·2	31·0	20·9	42·7	18·4	47·0	45·0	32·7	58·9	29·4	67·0
20·2	12·2	29·8	10·3	33·5	31·2	21·0	42·9	18·5	47·3	45·5	33·1	59·5	29·8	64·6
20·4	12·3	30·0	10·4	33·8	31·4	21·2	43·2	18·6	47·6	46·0	33·5	60·0	30·2	65·2
20·6	12·5	30·3	10·6	34·1	31·6	21·4	43·4	18·8	47·8	46·5	33·9	60·6	30·6	65·8
20·8	12·6	30·5	10·7	34·3	31·8	21·5	43·6	18·9	48·1	47·0	34·4	61·2	31·1	66·4
21·0	12·8	30·7	10·9	34·5	32·0	21·7	43·8	19·1	48·3	47·5	34·8	61·8	31·4	67·0
21·2	13·0	31·0	11·0	34·8	32·2	21·9	44·1	19·3	48·6	48·0	35·2	62·3	31·9	67·5
21·4	13·1	31·3	11·1	35·1	32·4	22·0	44·3	19·4	48·8	48·5	35·6	62·9	32·2	68·2
21·6	13·3	31·5	11·3	35·3	32·6	22·2	44·6	19·6	49·1	49·0	36·1	63·5	32·7	68·7
21·8	13·4	31·7	11·4	35·6	32·8	22·4	44·8	19·7	49·3	49·5	36·5	64·1	33·1	69·4
22·0	13·6	31·9	11·6	35·8	33·0	22·5	45·0	19·9	49·5	50·0	36·9	64·6	33·5	69·9
22·2	13·7	32·2	11·7	36·1	33·2	22·7	45·3	20·0	49·8	50·5	37·3	65·2	33·9	70·6
22·4	13·9	32·5	11·9	36·4	33·4	22·8	45·5	20·2	50·0	51·0	37·8	65·8	34·3	71·1
22·6	14·1	32·7	12·0	36·6	33·6	23·0	45·7	20·3	50·3	51·5	38·2	66·4	34·7	71·7
22·8	14·2	32·9	12·2	36·8	33·8	23·2	46·0	20·5	50·5	52·0	38·7	66·9	35·1	72·3
23·0	14·4	33·1	12·4	37·0	34·0	23·4	46·2	20·7	50·7	52·5	39·1	67·5	35·5	72·9
23·2	14·5	33·4	12·5	37·3	34·2	23·5	46·4	20·8	51·0	53·0	39·5	68·0	36·0	73·5
23·4	14·7	33·7	12·6	37·6	34·4	23·7	46·7	21·0	51·3	53·5	39·9	68·6	36·3	74·1
23·6	14·8	33·9	12·7	37·9	34·6	23·8	46·9	21·1	51·5	54·0	40·4	69·2	36·8	74·6
23·8	15·0	34·1	12·9	38·1	34·8	24·0	47·1	21·3	51·7	54·5	40·8	69·8	37	75
24·0	15·2	34·3	13·1	38·3	35·0	24·2	47·3	21·5	51·9	55·0	41·3	70·3	38	76
24·2	15·3	34·6	13·2	38·6	35·2	24·4	47·6	21·6	52·2	56	42·1	71·4	38	77
24·4	15·5	34·9	13·3	38·9	35·4	24·5	47·8	21·7	52·5	57	43·0	72·6	39	78
24·6	15·7	35·1	13·5	39·1	35·6	24·7	48·1	21·9	52·7	58	43·9	73·7	40	79
24·8	15·8	35·3	13·6	39·4	35·8	24·9	48·3	22·1	52·9	59	44·7	74·8	41	80
25·0	16·0	35·5	13·8	39·6	36·0	25·0	48·5	22·3	53·2	60	46	76	42	82
25·2	16·1	35·8	14·0	39·9	36·2	25·2	48·8	22·4	53·4	61	46	77	43	83
25·4	16·3	36·1	14·1	40·1	36·4	25·3	49·0	22·5	53·7	62	47	78	43	84
25·6	16·5	36·3	14·2	40·4	36·6	25·5	49·2	22·7	53·9	63	48	79	44	85
25·8	16·6	36·5	14·4	40·6	36·8	25·7	49·5	22·9	54·2	64	49	80	45	86
26·0	16·8	36·7	14·6	40·8	37·0	25·9	49·7	23·0	54·4	65	50	82	46	87
26·2	17·0	37·0	14·7	41·1	37·2	26·0	49·9	23·2	54·6	66	51	83	47	89
26·4	17·1	37·3	14·8	41·4	37·4	26·2	50·2	23·3	54·9	67	52	84	48	90
26·6	17·3	37·5	15·0	41·6	37·6	26·4	50·4	23·5	55·1	68	53	85	48	91
26·8	17·4	37·7	15·1	41·9	37·8	26·5	50·6	23·6	55·4	69	54	86	49	92
27·0	17·6	37·9	15·3	42·1	38·0	26·7	50·8	23·8	55·6	70	54	87	50	93
27·2	17·8	38·2	15·5	42·4	38·2	26·9	51·1	24·0	55·9	71	55	88	51	94
27·4	17·9	38·4	15·6	42·6	38·4	27·0	51·3	24·1	56·1	72	56	89	52	96
27·6	18·1	38·7	15·7	42·9	38·6	27·2	51·6	24·3	56·4	73	57	91	53	97
27·8	18·3	38·9	15·9	43·1	38·8	27·4	51·8	24·4	56·6	74	58	92	54	98
28·0	18·4	39·1	16·1	43·3	39·0	27·6	52·0	24·6	56·8	75	59	93	54	99
28·2	18·6	39·4	16·2	43·6	39·2	27·7	52·2	24·8	57·1	76	60	94	55	100
28·4	18·7	39·6	16·3	43·9	39·4	27·9	52·5	24·9	57·3	77	61	95	56	101
28·6	18·9	39·9	16·5	44·1	39·6	28·0	52·7	25·1	57·6	78	62	96	57	102
28·8	19·1	40·1	16·7	44·4	39·8	28·2	52·9	25·2	57·8	79	62	97	58	104
29·0	19·3	40·3	16·8	44·6	40·0	28·4	53·1	25·4	58·0	80	63	98	69	105
29·2	19·4	40·6	17·0	44·9	40·5	28·8	53·8	25·8	58·6	81	64	99	60	106
29·4	19·6	40·8	17·1	45·1	41·0	29·3	54·3	26·2	59·2	82	65	101	60	107
29·6	19·7	41·0	17·2	45·4	41·5	29·6	54·9	26·6	59·8	83	66	102	61	108
29·9	19·9	41·3	17·4	45·6	42·0	30·1	55·5	27·0	60·4	84	67	103	62	109
30·0	20·1	41·5	17·6	45·8	42·5	30·5	56·1	27·4	61·0	85	68	104	63	110

$X+Y$	$\alpha=0.95$		$\alpha=0.99$		$X+Y$	$\alpha=0.95$		$\alpha=0.99$		$X+Y$	$\alpha=0.95$		$\alpha=0.99$	
86	69	105	64	112	141	119	165	112	173	196	169	224	162	234
87	70	106	65	113	142	119	166	113	174	197	170	225	163	235
88	70	107	66	114	143	120	167	114	176	198	171	226	163	236
89	71	108	66	115	144	121	168	115	177	199	172	227	164	237
90	72	109	67	116	145	122	169	116	178	200	173	228	165	238
91	73	110	68	117	146	123	170	117	179	201	174	230	166	239
92	74	112	69	118	147	124	172	117	180	202	175	231	167	240
93	75	113	70	120	148	125	173	118	181	203	176	232	168	241
94	76	114	71	121	149	126	174	119	182	204	177	233	169	242
95	77	115	72	122	150	127	175	120	183	205	178	234	170	244
96	78	116	72	123	151	128	176	121	184	206	179	235	171	245
97	78	117	73	124	152	129	177	122	185	207	180	236	172	246
98	79	118	74	125	153	130	178	123	187	208	181	237	173	247
99	80	119	75	126	154	130	179	124	188	209	181	238	173	248
100	81	120	76	127	155	131	180	125	189	210	182	239	174	249
101	82	121	77	129	156	132	181	126	190	211	183	240	175	250
102	83	123	78	130	157	133	182	126	191	212	184	241	176	251
103	84	124	79	131	158	134	183	127	192	213	185	242	177	252
104	85	125	79	132	159	135	184	128	193	214	186	243	178	253
105	86	126	80	133	160	136	186	129	194	215	187	245	179	254
106	87	127	81	134	161	137	187	130	195	216	188	246	180	256
107	88	128	82	135	162	138	188	131	196	217	189	247	181	257
108	88	129	83	136	163	139	189	132	198	218	190	248	182	258
109	89	130	84	138	164	140	190	133	199	219	191	249	183	259
110	90	131	85	139	165	141	191	134	200	220	192	250	184	260
111	91	132	86	140	166	142	192	135	201	221	193	251	184	261
112	92	134	86	141	167	142	193	135	202	222	194	252	185	262
113	93	135	87	142	168	143	194	136	203	223	195	253	186	263
114	94	136	88	143	169	144	195	137	204	224	195	254	187	264
115	95	137	89	144	170	145	196	138	205	225	196	255	188	265
116	96	138	90	145	171	146	197	139	206	226	197	256	189	266
117	97	139	91	147	172	147	198	140	207	227	198	257	190	268
118	98	140	92	148	173	148	200	141	209	228	199	258	191	269
119	98	141	93	149	174	149	201	142	210	229	200	259	192	270
120	99	142	93	150	175	150	202	143	211	230	201	260	193	271
121	100	143	94	151	176	151	203	144	212	231	202	262	194	272
122	101	144	95	152	177	152	204	144	213	232	203	263	194	273
123	102	146	96	153	178	153	205	145	214	233	204	264	195	274
124	103	147	97	154	179	154	206	146	215	234	205	265	196	275
125	104	148	98	156	180	154	207	147	216	235	206	266	197	276
126	105	149	99	157	181	155	208	148	217	236	207	267	198	277
127	106	150	100	158	182	156	209	149	218	237	208	268	199	278
128	107	151	101	159	183	157	210	150	220	238	209	269	200	279
129	108	152	101	160	184	158	211	151	212	239	209	270	201	281
130	108	153	102	161	185	159	212	152	222	240	210	271	202	282
131	109	154	103	162	186	160	213	153	223	241	211	272	203	283
132	110	155	104	163	187	161	215	153	225	242	212	273	204	284
133	111	156	105	164	188	162	216	154	225	243	213	274	205	285
134	112	157	106	166	189	163	217	155	226	244	214	275	205	286
135	113	159	107	167	190	164	218	156	227	245	215	276	206	287
136	114	160	108	168	191	165	219	157	228	246	216	278	207	288
137	115	161	109	169	192	166	220	158	229	247	217	279	208	289
138	116	162	109	170	193	167	221	159	230	248	218	280	209	290
139	117	163	110	171	194	167	222	160	232	249	219	281	210	291
140	118	164	111	172	195	168	223	161	233	250	220	282	211	292

Note: the pairs of figures under each α-heading are lower and upper confidence limits for λ.

Upper percentage points of a substitute F-ratio using ranges

By K. C. SREEDHARAN PILLAI† and ANGELES R. BUENAVENTURA‡
The Statistical Center, University of the Philippines

Let $F' = w_1/w_2$ where w_1 and w_2 are ranges of two independent samples of size n_1 and n_2, respectively, taken from normal populations having the same standard deviation. Table 2 gives the upper 5 and 1 % points of F', a substitute F-ratio, for values of n_1, n_2 varying from 2 to 10. This extends the work of Pillai (1951, 1957) who gave upper 5 and 1 % points of F' for n_1, n_2 ranging up to 8. The present work has also been directed towards correcting the values of F' which can be derived from the table of Link (1950) by taking reciprocals. His values for smaller sample sizes are not accurate enough for practical use.

Tabulation of the percentage points was made by extending Pillai's methods. First, his approximation to the distribution of the semi-range, W, as a series of gamma functions (Pillai, 1948, 1950, 1951) was improved by considering more terms in the expansion and the coefficients of the first few terms were tabulated (Buenaventura, 1959) for values of sample size from 3 to 12 instead of 3 to 8 as given by Pillai (1952). The distribution of the semi-range derived in this manner is given by

$$f(W) = W^{n-2} e^{-\frac{1}{8}(n+4)W^2} D, \tag{1}$$

where $$D = C_0^{(n)} + C_1^{(n)} W^2 + C_2^{(n)} W^4 + \ldots,$$

and n is the size of the sample, taken from a normal population with unit standard deviation, from which the semi-range W was obtained. The $C_i^{(n)}$ coefficients for values for n ranging from 3 to 12 are given in Table 1.

Using $f(W)$ in (1), the distribution of F' was obtained following Pillai (1951) in the form

$$f(F') = \frac{1}{2} \sum_{r=0}^{\infty} \sum_{j=0}^{r} \Gamma\{\tfrac{1}{2}(n_1+n_2)+r-1\} C_{r-j}^{(n_1)} C_j^{(n_2)} (F')^{n_1+2(r-j-1)} \{\tfrac{1}{8}[n_2+4+(n_1+4)F'^2]\}^{-\frac{1}{2}(n_1+n_2)-r+1}. \tag{2}$$

Percentage points presented in Table 2 were based on series (2). The incomplete beta functions involved in the cumulative distribution function of F' derivable from (2) were mostly integrated by parts in order to have sufficient accuracy. For $n_2 = 2$, the values in Table 2 can be computed by dividing by $\sqrt{2}$, the corresponding percentage points of the studentized range (Pillai, 1957; Harter, Clemm & Guthrie, 1959).

The authors wish to acknowledge the facilities offered by the Statistical Center, University of the Philippines, in the preparation of this paper.

References

Buenaventura, A. R. (1959). *On a Substitute F-ratio Using Ranges.* M.A.S. Thesis submitted to the Graduate School, University of the Philippines.

Harter, H. L., Clemm, D. S. & Guthrie, E. H. (1959). *The Probability Integral of the Range and of the Studentized Range.* Ohio: Wright Air Development Center. Tech. Report No. 58-484, **2**.

Link, R. F. (1950). The sampling distribution of the ratio of two ranges from independent samples. *Ann. Math. Statist.* **21**, 112–6.

Pillai, K. C. S. (1948). A note on ordered samples. *Sankhyā*, **8**, 375–80.

Pillai, K. C. S. (1950). On the distributions of mid-range and semi-range in samples from a normal population. *Ann. Math. Statist*, **21**, 100–5.

Pillai, K. C. S. (1951). Some notes on ordered samples from a normal population. *Sankhyā*, **11**, 23–8.

Pillai, K. C. S. (1952). On the distribution of 'Studentized' range. *Biometrika*, **39**, 194–5.

Pillai, K. C. S. (1957). *Concise Tables for Statisticians.* Manila: The Statistical Center.

† United Nations Senior Adviser in Mathematical Statistics and Visiting Professor of Statistics. Now with United Nations, New York.

‡ Research Instructor.

Table 1. *Values of $C_i^{(n)}$ coefficients for $n = 3, 4, \ldots, 12$*

n	$C_0^{(n)}$	$C_1^{(n)}$	$C_2^{(n)}$	$C_3^{(n)}$
3	2·2053 15582	0·1228 72039	0·01417 55429	$0\cdot0^{3}4572\ 57432$
4	3·0476 94522	0·2542 84571	0·05915 73835	$0\cdot0^{2}1347\ 65185$
5	3·6249 76870	0·3627 24248	0·1195 48715	$0\cdot0^{2}4726\ 25527$
6	3·96046 2787	0·4402 10583	0·1833 01703	$0\cdot0^{2}9909\ 48370$
7	4·0958 17895	0·4877 08428	0·2426 00723	0·01585 23602
8	4·0758 99468	0·5095 65175	0·2929 29359	0·02170 90301
9	3·9421 37188	0·5110 72548	0·3321 43307	0·02691 69907
10	3·7299 50454	0·4973 65581	0·3597 20048	0·03116 04640
11	3·4681 42690	0·4729 56308	0·3761 83911	0·03430 93722
12	3·1792 44908	0·4415 81756	0·3826 80232	0·03636 18641

n	$C_4^{(n)}$	$C_5^{(n)}$	$C_6^{(n)}$	$C_7^{(n)}$
3	$0\cdot0^{4}2189\ 56371$	$0\cdot0^{6}6973\ 23557$	$0\cdot0^{8}1046\ 91042$	$0\cdot0^{8}2219\ 29693$
4	$0\cdot0^{3}4576\ 33423$	$-0\cdot0^{5}7798\ 29201$	$0\cdot0^{5}1562\ 44912$	$0\cdot0^{6}1801\ 46225$
5	$0\cdot0^{2}1430\ 01981$	$0\cdot0^{5}7386\ 93546$	$0\cdot0^{5}3400\ 48316$	$0\cdot0^{5}2766\ 52948$
6	$0\cdot0^{3}3170\ 05571$	$0\cdot0^{4}5244\ 90681$	$0\cdot0^{4}1031\ 01293$	$0\cdot0^{4}1046\ 34952$
7	$0\cdot0^{2}5610\ 12308$	$0\cdot0^{3}1483\ 22530$	$0\cdot0^{4}2932\ 30434$	$0\cdot0^{4}2359\ 09785$
8	$0\cdot0^{2}8533\ 62668$	$0\cdot0^{3}2995\ 52913$	$0\cdot0^{4}6788\ 12458$	$0\cdot0^{4}4114\ 14903$
9	0·01168 44731	$0\cdot0^{3}4963\ 15277$	$0\cdot0^{3}1335\ 27171$	$0\cdot0^{4}5984\ 07437$
10	0·01482 42995	$0\cdot0^{3}7217\ 10268$	$0\cdot0^{3}2268\ 90958$	$0\cdot0^{4}7809\ 74046$
11	0·01775 83304	$0\cdot0^{3}9573\ 63192$	$0\cdot0^{3}3447\ 66736$	$0\cdot0^{4}9470\ 27794$
12	0·02034 36661	$0\cdot0^{2}1186\ 71181$	$0\cdot0^{3}4810\ 41785$	$0\cdot0^{3}1090\ 05391$

Table 2. *Percentage points of the distribution of $F' = w_1/w_2$**

n_2 \ n_1	2	3	4	5	6	7	8	9	10
				Upper 5 % points					
2	12·71	19·08	23·2	26·2	28·6	30·5	32·1	33·5	34·7
3	3·19	4·37	5·13	5·72	6·16	6·53	6·85	7·12	7·33
4	2·03	2·66	3·08	3·38	3·62	3·84	4·00	4·14	4·26
5	1·60	2·05	2·35	2·57	2·75	2·89	3·00	3·11	3·19
6	1·38	1·74	1·99	2·17	2·31	2·42	2·52	2·61	2·69
7	1·24	1·57	1·77	1·92	2·04	2·13	2·21	2·28	2·34
8	1·15	1·43	1·61	1·75	1·86	1·94	2·01	2·08	2·13
9	1·09	1·33	1·49	1·62	1·72	1·79	1·86	1·92	1·96
10	1·05	1·26	1·42	1·54	1·63	1·69	1·76	1·82	1·85
				Upper 1 % points					
2	63·66	95·49	116·1	131	143	153	161	168	174
3	7·37	10·00	11·64	12·97	13·96	14·79	15·52	16·13	16·60
4	3·73	4·79	5·50	6·01	6·44	6·80	7·09	7·31	7·51
5	2·66	3·33	3·75	4·09	4·36	4·57	4·73	4·89	5·00
6	2·17	2·66	2·98	3·23	3·42	3·58	3·71	3·81	3·88
7	1·89	2·29	2·57	2·75	2·90	3·03	3·13	3·24	3·33
8	1·70	2·05	2·27	2·44	2·55	2·67	2·76	2·84	2·91
9	1·57	1·89	2·07	2·22	2·32	2·43	2·50	2·56	2·63
10	1·47	1·77	1·92	2·06	2·16	2·26	2·33	2·38	2·44

* w_1 denotes the larger range. For a test of equality of two population variances, the 5 % and 1 % levels of this table are the 10 % and 2 % levels, respectively, of a two-tailed test.

On Durbin's formula for the limiting generalized variance of a sample of consecutive observations from a moving-average process

BY A. M. WALKER

Statistical Laboratory, University of Cambridge

1. Let $x_1, x_2, \ldots, x_n$ be consecutive observations from the hth order moving-average process $\{x_t\}$ defined by

$$x_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_h \epsilon_{t-h},$$

where $\{\epsilon_t\}$ $(t = 0, \pm 1, \pm 2, \ldots)$ is a set of independent random variables having a common distribution with mean zero and variance σ^2. Let the variance-covariance matrix of $x_1, x_2, \ldots, x_n$ be $\sigma^2 V_n$, so that if $V_n = (v_{ij})$ $(i, j = 1, 2, \ldots, n)$,

$$v_{ij} = E(x_i x_{i+|i-j|})/\sigma^2 = \sum_{r=0}^{h} \beta_r \beta_{r+|i-j|},$$

where we define $\beta_0 = 1$ and $\beta_r = 0$ for $r > h$. Let the determinant $|V_n|$ be denoted by D_n, so that the generalized variance of the set of observations is $\sigma^2 D_n$. Durbin (1959) has shown that for $h = 1, 2$, $\lim_{n\to\infty} D_n$ exists provided that the roots of the equation $z^h + \beta_1 z^{h-1} + \ldots + \beta_h = 0$ each have modulus less than unity, and is equal to $\lim_{n\to\infty} D'_n$, where $\sigma^2 D'_n$ is the generalized variance of a similar set of observations from the autoregressive process which is the stationary solution of the equation

$$x_t + \beta_1 x_{t-1} + \ldots + \beta_h x_{t-h} = \epsilon_t.$$

Durbin conjectured that the result holds for any value of h and in this note we prove that his conjecture is correct.

A proof has also recently been given by Finch (1960). He makes use of the theory of Toeplitz forms, and is thereby able to obtain a generalization of Durbin's result. However, the present proof, while perhaps less elegant than Finch's, is entirely self-contained, depending only on elementary properties of determinants and matrices. It also shows that $\lim_{n\to\infty} D_n$ is equal to a determinant of order h whose elements are quadratic polynomials in $\beta_1, \beta_2, \ldots, \beta_h$, and this is more easily evaluated than the expression for the limit given by Finch, which is a function of the roots of the equation

$$z^h + \beta_1 z^{h-1} + \ldots + \beta_h = 0.$$

2. Define β_r to be zero for $r < 0$ as well as for $r > h$, and let B_n be the matrix with β_{j-i} as the element in row i and column j $(i, j = 1, 2, \ldots, n)$. Then taking $n > h$, we have

$$(B'_n B_n)_{ij} = \sum_{r=1}^{n} \beta_{i-r}\beta_{j-r} = \sum_{r=-\infty}^{\infty} \beta_{i-r}\beta_{j-r} = v_{ij}$$

whenever $i > h$ and $j > h$. Hence if $U_n = V_n - B'_n B_n$, we can write

$$U_n = \begin{pmatrix} U_h & 0 \\ 0 & 0 \end{pmatrix},$$

where

$$(U_h)_{ij} = \sum_{r=-(h-1)}^{0} \beta_{i-r}\beta_{j-r} = \sum_{r=0}^{h-1} \beta_{i+r}\beta_{j+r} \quad (i, j = 1, 2, \ldots, h).$$

Thus

$$V_n^{-1} B'_n B_n = I_n - \begin{pmatrix} V_n^{11} & V_n^{12} \\ V_n^{21} & V_n^{22} \end{pmatrix} \begin{pmatrix} U_h & 0 \\ 0 & 0 \end{pmatrix},$$

where V_n^{-1} is partitioned into submatrices in the same way as U_n, and I_n is the $n \times n$ unit matrix. Therefore

$$V_n^{-1} B'_n B_n = \begin{pmatrix} I_h - V_n^{11} U_h & 0 \\ V_n^{21} U_h & I_{n-h} \end{pmatrix},$$

and, as $|B_n| = 1$,

$$|V_n^{-1}| = |V_n^{-1} B'_n B_n| = |I_h - V_n^{11} U_h|. \tag{1}$$

We now require the following lemma.

LEMMA.

$$\lim_{n\to\infty} (V_n^{-1})_{ij} = \sum_{r=1}^{h} b^{ir} b^{jr},$$

where

$$B_n^{-1} = (b^{ij}) \quad (i, j = 1, 2, \ldots, n)$$

(b^{ij} does not depend on n because B_n, and therefore also B_n^{-1}, is an upper triangular matrix).

Proof. Let $E_n = V_n - C_n$, where $C_n = B_n B_n'$. Then $V_n(V_n^{-1} - C_n^{-1})C_n = -E_n$, so that

$$V_n^{-1} = C_n^{-1} - V_n^{-1} E_n C_n^{-1}.$$

Now
$$(E_n)_{ij} = \sum_{r=-\infty}^{\infty} \beta_{r-i}\beta_{r-j} - \sum_{r=1}^{n} \beta_{r-i}\beta_{r-j} = \sum_{r=n+1}^{\infty} \beta_{r-i}\beta_{r-j},$$

which is zero unless $|i-j| \leqslant h$ and either i or $j > n-h$, and therefore unless both i and $j > n-2h$. Thus, writing V_n^{ij} for $(V_n^{-1})_{ij}$,

$$V_n^{ij} - (C_n^{-1})_{ij} = \sum_{r,s=n-2k+1}^{n} V_n^{ir}(E_n)_{rs}(C_n^{-1})_{sj}. \tag{2}$$

Let $\sum_{i=0}^{\infty} \alpha_i \theta^i$ be the expansion of $(1+\beta_1\theta + \ldots + \beta_h\theta^h)^{-1}$ in powers of θ (which is certainly valid for $|\theta| \leqslant 1$). Then it is easy to verify that $b^{ij} = \alpha_{j-1}$ for $j \geqslant i$, and therefore, if we define α_i to be zero for $i < 0$,

$$(C_n^{-1})_{ij} = \sum_{r=1}^{n} \alpha_{r-i}\alpha_{r-j}.$$

Hence
$$\lim_{n\to\infty} (C_n^{-1})_{n-i,\,j} = \lim_{n\to\infty} \sum_{r=0}^{i} \alpha_r \alpha_{r+n-i+j} = 0.$$

Now clearly $\sup_n |(E_n)_{n-i+1,\,n-j+1}| < \infty$, where (i,j) is any fixed pair of positive integers, and the supremum is taken over all positive integral values of n. The right-hand side of (2) will therefore certainly tend to zero as $n \to \infty$, thus establishing the lemma, provided that $\sup_n |V_n^{i,\,n-j+1}| < \infty$. Now V_n and therefore V_n^{-1} is positive definite, so that

$$(V_n^{i,\,n-j+1})^2 < V_n^{ii} V_n^{n-j+1,\,n-j+1} = V_n^{n-i+1,\,n-i+1} V_n^{n-j+1,\,n-j+1}$$

(since V_n is not altered when the order of its rows and columns is reversed). Hence it will suffice to show that for any fixed i, $\sup_n V_n^{n-i+1,\,n-i+1} < \infty$. But $V_n^{n-i+1,\,n-i+1} = (1-R_{n,i}^2)^{-1}$, where $R_{n,i}$ is the multiple correlation coefficient between x_{n-i+1} and $(x_1, x_2, \ldots, x_{n-i}, x_{n-i+2}, \ldots, x_n)$ (see, for example, Cramér, 1946, p. 308). It is easily seen, by expressing $x_1, x_2, \ldots, x_n$ in terms of the ϵ_t, that the 'residual' variable defined by subtracting from x_{n-i+1} its linear regression on the remaining x's has a variance which is not less than unity. It follows that $\sup_n R_{n,i}^2 < 1$, and therefore $\sup_n V_n^{n-i+1,\,n-i+1} < \infty$.

Let $a_{ij} = \lim_{n\to\infty} V_n^{ij}$ and $A = (a_{ij})$ $(i,j = 1, 2, \ldots, h)$. Then if $(a^{ij}) = A^{-1}$, it follows from the lemma that

$$a^{ij} = \sum_{r=1}^{h} \beta_{r-i}\beta_{r-j}, \quad \text{i.e.} \quad A^{-1} = B_h B_h'.$$

Therefore from the equation (1)

$$\lim_{n\to\infty} |V_n^{-1}| = |I_h - A U_h| = |A|\,|A^{-1} - U_h|.$$

Also $|A| = 1$, so that finally
$$\lim_{n\to\infty} |V_n^{-1}| = \lim_{n\to\infty} D_n^{-1} = |B_h B_h' - U_h|.$$

3. We now show that $\lim_{n\to\infty} D_n' = |B_h B_h' - U_h|^{-1}$.

First we note that D_n' is constant for $n \geqslant h$; this follows from the fact that when $\{x_t\}$ is an autoregressive process generated by the equation

$$x_t + \beta_1 x_{t-1} + \ldots + \beta_h x_{t-h} = \epsilon_t,$$

$$\left[\sum_{t=h+1}^{n} (x_t + \beta_1 x_{t-1} + \ldots + \beta_h x_{t-h})^2/\sigma^2 + \text{a quadratic form in } x_1, x_2, \ldots, x_h\right]$$

is the sum of squares of n uncorrelated variables with unit variances [compare Siddiqui, 1958, p. 586, equation (7)]. It will therefore suffice to prove that $D_h' = |B_h B_h' - U_h|^{-1}$.

Let W_n denote the variance-covariance matrix of n consecutive observations from the above autoregressive process, and X_n' denote the row vector $(x_1, x_2, \ldots, x_n)$. Then we have

$$X_n' W_n^{-1} X_n = X_h' W_h^{-1} X_h + \sum_{t=h+1}^{n} \left(\sum_{i=0}^{h} \beta_i x_{t-i}\right) \Big/ \sigma^2 \quad (n > h)$$

as an identity in $(x_1, x_2, \ldots, x_n)$ (Siddiqui, 1958, equation (8)), and Siddiqui shows how this may be used to obtain the elements of W_n^{-1} for $n > 2h$. His argument (which is also given by Champernowne, 1948, p. 206) is in fact valid for $n = 2h$, and using the result

$$\sigma^2 (W_{2h}^{-1})_{ij} = \sum_{r=1}^{h} \beta_{i-r}\beta_{j-r} \quad \text{if} \quad i \leqslant h, \quad j \leqslant h,$$

we obtain, on putting $n = 2h$ in the above identity,

$$\sigma^2(W_h^{-1})_{ij} = \sum_{r=1}^{h} \beta_{i-r}\beta_{j-r} - \sum_{r=h+1}^{2h} \beta_{r-i}\beta_{r-j}$$

(compare Durbin, 1959, p. 311, equation (19)). From this it is easily verified that

$$\sigma^2(W_h^{-1})_{ij} = (B_h B_h' - U_h)_{ij},$$

$$\left[\text{for } i \geqslant j, \text{ each is equal to } \left(\sum_{r=0}^{j-1} - \sum_{r=h+1-i}^{h+j-i}\right)\beta_r\beta_{r+i-j}\right]$$

so that $(D_h')^{-1} = |B_h B_h' - U_h|$. Hence we have established that $\lim_{n\to\infty} D_n' = \lim_{n\to\infty} D_n$.

REFERENCES

CHAMPERNOWNE, D. G. (1948). Sampling theory applied to autoregressive sequences. *J. R. Statist. Soc.* B, **10**, 204–31.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

DURBIN, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika*, **46**, 306–16.

FINCH, P. D. (1960). On the covariance determinants of moving-average and autoregressive models. *Biometrika*, **47**, 194–6.

SIDDIQUI, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Statist.* **29**, 585–8.

The central sampling moments of the mean in samples from a finite population (Aty's formulae and Madow's central limit)

BY D. E. BARTON AND F. N. DAVID†

University College London

We consider a sample $x_1, x_2, \ldots, x_n$ randomly drawn without replacement from a finite population $\mathscr{F} \equiv (X_1, X_2, \ldots, X_N)$. The polykays, as defined by Tukey (1950), of this population may be written as

$$K_{p_1\pi_1 \ldots p_\lambda\pi_\lambda}.$$

Aty (1954), gave the first eight central moments of the sample mean, k_1, in terms of these polykays, and of constants

$$A_r = \sum_{i=1}^{r-1} \alpha^i \left(-\frac{1}{N}\right)^{r-1-i} \qquad (r = 1, 2, \ldots, 8),$$

where

$$\alpha = \left(\frac{1}{n} - \frac{1}{N}\right),$$

but did not establish the general relationship,

$$M_{\mathscr{F}}(1^r) = \sum_{P} \frac{r!\,(\pi-1)!}{(p_1!)^{\pi_1} \ldots (p_\lambda!)^{\pi_\lambda} \pi_1! \ldots \pi_\lambda!} A_{p_1}^{\pi_1} \ldots A_{p_\lambda}^{\pi_\lambda} K_{p_1\pi_1 \ldots p_\lambda\pi_\lambda}. \tag{1}$$

In this expression the suffix $\mathscr{F}$ denotes 'in sampling from $\mathscr{F}$', the summation is over all partitions $(p_1^{\pi_1} \ldots p_\lambda^{\pi_\lambda})$ of r excluding those with unit parts, λ is the number of *distinct* parts and

$$\pi = \pi_1 + \pi_2 + \ldots + \pi_\lambda,$$

is the *number* of parts. If the $\{X_i\}$ are themselves a sample from an infinite population which has finite cumulants $\kappa_1, \kappa_2, \ldots$ then, since A_r tends to $n^{-(r-1)}$ as N tends to infinity, the right-hand side of (1) gives the usual expression for the moments of k_1 in independent sampling, on taking expectations and letting N tend to infinity. The numerical coefficients are thus well tabulated.

We have found the general validation of (1) to be necessary for theoretical purposes (Barton, David & Fix, 1960) in spite of the fact that in practice Aty's results are sufficient. The general formulae simplify Madow's (1948) central limit theorem for finite populations and slightly extend it.

† With the partial support of the Office of Naval Research while at the University of California, Berkeley.

Proof of Aty's formulae

Since $M_{\mathscr{F}}(1^r)$ is necessarily a polynomial, homogeneous and of weight r, in the moments of $\mathscr{F}$ and since products of these are uniquely expressible as linear sums of the polykays of weight r (cf. Wishart, 1952), then $M_{\mathscr{F}}(1^r)$ is uniquely expressible as a linear sum of polykays of weight r. Considering $\mathscr{F}$, formally, to be a sample of N from an infinite population $\mathscr{I}$ whose cumulants are $\kappa_1, \kappa_2, \ldots$ we have from the definition of the polykays

$$\mathscr{E}_{\mathscr{I}}(K_{p_1^{\pi_1}\ldots p_\lambda^{\pi_\lambda}}) = \kappa_{p_1}^{\pi_1}\ldots\kappa_{p_\lambda}^{\pi_\lambda},$$

and accordingly it is sufficient to show that

$$\mathscr{E}_{\mathscr{I}}(M_{\mathscr{F}}(1^r)) = \sum_P \frac{r!\,(\pi-1)!}{(p_1!)^{\pi_1}\ldots(p_\lambda!)^{\pi_\lambda}\pi_1!\ldots\pi_\lambda!} A_{p_1}^{\pi_1}\ldots A_{p_\lambda}^{\pi_\lambda}\kappa_{p_1}^{\pi_1}\ldots\kappa_{p_\lambda}^{\pi_\lambda}. \tag{2}$$

For, any other coefficients would imply a generally persisting polynomial identity between $\kappa_1, \ldots, \kappa_r$, and this by the general theory of moments does not exist. The relation (2) is easily seen to be true. If $y_1, \ldots, y_{N-n}$ denote the residue of $\mathscr{F}$ after $x_1, \ldots, x_n$ have been removed,

$$\bar{x} - K_1 = \left(\frac{N-n}{N}\right)(\bar{x}-\bar{y}) = \frac{N-n}{N}((\bar{x}-\kappa_1)-(\bar{y}-\kappa_1)). \tag{3}$$

Accordingly the rth cumulant of $(\bar{x}-K_1)$ in sampling from $\mathscr{I}$ is

$$\left(\frac{N-n}{N}\right)^r [n^{-(r-1)} + (-1)^r (N-n)^{-(r-1)}]\kappa_r = A_r\kappa_r. \tag{4}$$

Putting (4) into the general expression for moments in terms of cumulants we have (2) which is sufficient to prove (1).

It may be noted that, while for the purposes of the present proof $\mathscr{F}$ has been assumed to be a sample from an infinite population $\mathscr{I}$, since (1) is a purely algebraic expression for $M_{\mathscr{F}}(1^r)$ in terms of $X_1, \ldots, X_N$, it may be used when the finite population $\mathscr{F}$ does not necessarily have any such provenance.

Madow's central limit theorem

We consider here the behaviour of

$$u = (k_1 - K_1)\sqrt{n}/\sqrt{K_2}$$

under the double limit, $N \to \infty$, $n \to \infty$ with

$$\frac{n}{N} = p + O\left(\frac{1}{N}\right) \quad (0 < p < 1),$$

holding for fixed p. The behaviour of the polykays is to be specified. In any event we have

$$\mathscr{E}(u) = 0, \quad \mathscr{E}(u^2) \to 1-p.$$

When each of the $\{K_r/K_1^{\frac{1}{2}r}\}$ is bounded as $N \to \infty$, then it is seen from the form of the relation between the general polykays and those of single suffix that

$$K_2^{-\frac{1}{2}r} K_{p_1^{\pi_1}\ldots p_\lambda^{\pi_\lambda}} \sim K_2^{-\frac{1}{2}r} K_{p_1}^{\pi_1}\ldots K_{p_\lambda}^{\pi_\lambda}$$

is also bounded. Since

$$A_r = O(N^{1-r})$$

it follows that the rth moment of u, i.e.

$$\left(\frac{n}{K_2}\right)^{\frac{1}{2}r} M_{\mathscr{F}}(1^r) \to \begin{cases} 0, & r \text{ odd}, \\ (r-1)(r-3)\ldots 1(1-p)^{\frac{1}{2}r}, & r \text{ even}. \end{cases}$$

Since the normal distribution is uniquely determined by its moments, by the Fréchét–Shohat Second Limit theorem u tends to be normally distributed with zero mean and variance $(1-p)$. It will be noticed that we have required the K_r to be bounded but not necessarily uniformly as with Madow.

The same normal limit also holds if the $(X_1, \ldots, X_N)$ are regarded as a sample from one of a set of populations which satisfy the conditions of a central limit theorem, by virtue of (3).

Moments of k_2

The corresponding relation to (3) is

$$k_2 - K_2 = \left(\frac{N-n}{N-1}\right)k_2 - \left(\frac{N-n-1}{N-1}\right)k_2' - \frac{n(N-n)}{N(N-1)}(\bar{x}-\bar{y})^2,$$

where k_2 is the second k statistic of $x_1, x_2, \ldots, x_n$, and k_2' that of $y_1, y_2, \ldots, y_{N-n}$. This may alternatively be written as

$$\left(\frac{N-n}{N-1}\right)k_2 - \left(\frac{N-n-1}{N-1}\right)k_2' = k_2 - K_2 + \frac{nN}{(N-1)(N-n)}(k_1-K_1)^2.$$

The general form of the moments of k_2 in $\mathscr{F}$ is thus deducible from the moments of k_2 and k_1 in $\mathscr{S}$ jointly, but not from those of k_2 alone. The same is true of the joint moments of k_1 and k_2 in $\mathscr{F}$. The relationship is useful in that it enables us to exhibit these moments as rational functions of n with coefficients depending on the polykays.

References

Abdel-Aty, S. H. (1954). Tables of generalized k-statistics. *Biometrika*, **41**, 253–60.

Barton, D. E., David, F. N. & Fix, Evelyn (1960). The polykays of the natural numbers. *Biometrika*, **47**, 53–9.

Madow, W. G. (1948). On the limiting distributions of estimates based on samples from finite universes. *Ann. Math. Statist.* **19**, 535–45.

Tukey, J. W. (1950). Some sampling simplified. *J. Amer. Statist. Ass.* **45**, 501–19.

Wishart, J. (1952). Moment coefficients of the k-statistics in samples from a finite population. *Biometrika*, **39**, 1–13.

A note on the quadrivariate normal integral

By MAN MOHAN SONDHI

Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India

In this paper we derive an approximate closed expression for the function

$$P_4(\rho) = \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty p(x_1, x_2, x_3, x_4, \rho)\, dx_1 dx_2 dx_3 dx_4, \tag{1}$$

where $p(x_1, x_2, x_3, x_4, \rho)$ is the 4-dimensional normal density function with all off-diagonal elements of the associated correlation matrix equal to ρ. Thus $P_4(\rho)$ is the probability that all the four random variables obeying such a distribution are simultaneously positive. To the author's knowledge, no one has yet succeeded in expressing $P_4(\rho)$ in an exact closed form. McFadden (1956) has recently derived an expression which gives $P_4(\rho)$ accurate to 5 decimal places for $0 \leqslant \rho \leqslant \frac{1}{2}$, but the accuracy falls off rapidly for larger ρ. The expression derived in this paper (see equation (5) or (7)) matches the simplicity of McFadden's expression, and has 5 decimal place accuracy for $\frac{1}{2} \leqslant \rho \leqslant 1$. By means of the relation†

$$P_4\left(\frac{-\rho}{1+2\rho}\right) + P_4(\rho) = \frac{1}{8} + \frac{3}{4\pi}\left[\sin^{-1}\rho - \sin^{-1}\left(\frac{\rho}{1+2\rho}\right)\right] + \frac{3}{2\pi^2}\sin^{-1}\rho \sin^{-1}\left(\frac{\rho}{1+2\rho}\right) \tag{2}$$

which is proved in the appendix, $P_4(\rho)$ may be obtained in the range $-\frac{1}{3} \leqslant \rho \leqslant 0$ as well. Thus a convenient and sufficiently accurate representation of $P_4(\rho)$ is now made possible for the entire permissible range of ρ $(-\frac{1}{3} \leqslant \rho \leqslant 1)$.

To derive the expression mentioned, we first note that

$$P_4(\rho) = \frac{1}{16} + \frac{3}{4\pi}\sin^{-1}\rho + \frac{3}{2\pi^2}\int_0^{\sin^{-1}\rho} \operatorname{cosec}^{-1}(2 + \operatorname{cosec}\lambda)\, d\lambda. \tag{3}$$

This result can be proved, for example, by the use of an elegant theorem due to Price (1959); however, the proof is not detailed here. For convenience let

$$g(\rho) = \int_0^{\sin^{-1}\rho} \operatorname{cosec}^{-1}(2 + \operatorname{cosec}\lambda)\, d\lambda. \tag{4}$$

Now, intuitively, $P_4(1) = \frac{1}{2}$, as $\rho = 1$ implies perfect correlation of all the random variables. This fact may also be proved by substituting $\rho = 1$ in (2), and noting that $P_4(-\frac{1}{3}) = 0$. (In fact for the n-variate case $P_n(-1/(n-1)) = 0$.) Using the value of $P_4(1)$ we get from (3)

$$g(1) = \pi^2/24.$$

Once $g(1)$ is known, we may now expand $g(\rho)$ as a Taylor series about $\rho = 1$, since the derivatives of $g(\rho)$ with respect to $\sin^{-1}(\rho)$ are all easily obtained. On carrying out this expansion to eleven terms (of which five vanish) and substituting in (3) we get

$$P_4(\rho) \simeq \frac{1}{2} - \frac{3\phi}{2\pi^2}(\tfrac{1}{2}\pi + \sin^{-1}\tfrac{1}{3}) + \frac{3\phi^3}{\pi^2\sqrt{8}}(\tfrac{1}{3^6} + a\phi^2 + b\phi^4 + c\phi^6), \tag{5}$$

where

$$\phi = \cos^{-1}\rho, \quad a = \frac{1}{5!}\frac{23}{48}, \quad b = \frac{1}{7!}\frac{3727}{1152}, \quad c = \frac{1}{9!}\frac{3320309}{82944}. \tag{6}$$

Equation (5) is the desired expression.

† This relation has been derived independently by J. A. McFadden.

It may be noted that since $P_4(\frac{1}{2}) = 0{\cdot}2$, $g(\frac{1}{2}) = \frac{1}{120}\pi^2$, and hence $g(\rho)$ could be expanded around $\rho = \frac{1}{2}$. Although this might increase the range of validity of the expression for the same accuracy, the computation of the expansion coefficients around $\rho = \frac{1}{2}$ is extremely cumbersome.

A simple first-order Shanks transformation (1955) applied to the terms $a\phi^2 + b\phi^4 + c\phi^6$ in (5) gives

$$P_4(\rho) \simeq \frac{1}{2} - \frac{3\phi}{2\pi^2}(\tfrac{1}{2}\pi + \sin^{-1}\tfrac{1}{3}) + \frac{3\phi^3}{\pi^2\sqrt{8}}\left(\tfrac{1}{360} + \phi^2\frac{(ac-b^2)\,\phi^2 - ab}{c\phi^2 - b}\right), \tag{7}$$

where ϕ, a, b, c are as given in (6).

For checking the accuracy of (5) and (7), $P_4(\rho)$ was calculated by Moran's (1956) summation formula with an estimated accuracy to 8 decimal places. The comparison is shown in Table 1. It is seen that (7) is accurate to at least 5 decimal places throughout, and to at least 7 decimal places for $\rho \geqslant 0{\cdot}8$.

Table 1

	$P(\rho)$		
ρ	Eq. (5)	Eq. (7)	Moran
0·5	0·1999 9452	0·1999 9841	0·1999 9991
0·6	0·2334 5168	0·2334 5266	0·2334 5234
0·7	0·2706 8513	0·2706 8530	0·2706 8540
0·8	0·3139 8525	0·3139 8543	0·3139 8544
0·9	0·3693 1236	0·3693 1236	0·3693 1236

The simplicity of the present method compared to Moran's is evident from the following example. To obtain 5 figure accuracy (which is the least given by the present method) by Moran's method one has to sum at least 12 terms for $\rho = 0{\cdot}5$ and at least 26 terms for $\rho = 0{\cdot}9$. Moreover, the computation of each of these terms requires the values of two functions, which have to be obtained by complicated interpolation from tables. In contrast, (7) requires only one value of $\cos^{-1}\rho$ from tables, and far fewer terms have to be computed.

The first term of Shanks' second-order transformation was also computed, but it overcorrects the value at $\rho = \frac{1}{2}$ to 0·200085. The attempt was not carried to more terms or higher-order transformations.

The author acknowledges the helpful suggestions of Prof. J. A. McFadden. The financial assistance granted by the Council of Scientific and Industrial Research, India, is also gratefully acknowledged.

Appendix

To prove (2) it suffices to show that

$$g(\rho) + g\left(\frac{-\rho}{1+2\rho}\right) = \sin^{-1}\rho\,\sin^{-1}\left(\frac{\rho}{1+2\rho}\right). \tag{8}$$

Now
$$g(\rho) = \int_0^{\sin^{-1}\rho} \operatorname{cosec}^{-1}(2 + \operatorname{cosec}\lambda)\,d\lambda$$
$$= \int_0^{\rho} \sin^{-1}\left(\frac{\lambda}{1+2\lambda}\right)\frac{d}{d\lambda}\sin^{-1}\lambda\,d\lambda$$
$$= \sin^{-1}\rho\,\sin^{-1}\left(\frac{\rho}{1+2\rho}\right) - \int_0^{\rho}\sin^{-1}\lambda\,\frac{d}{d\lambda}\sin^{-1}\left(\frac{\lambda}{1+2\lambda}\right)d\lambda. \tag{9}$$

Also
$$g\left(\frac{-\rho}{1+2\rho}\right) = \int_0^{-\rho/(1+2\rho)} \sin^{-1}\left(\frac{\lambda}{1+2\lambda}\right)\frac{d}{d\lambda}\sin^{-1}\lambda\,d\lambda$$
$$= \int_0^{\rho}\sin^{-1}\mu\,\frac{d}{d\mu}\sin^{-1}\left(\frac{\mu}{1+2\mu}\right)d\mu, \tag{10}$$

where the change of variable of integration $\mu = -\lambda/(1+2\lambda)$ is made. Adding (9) and (10), we get (8) and hence (2).

REFERENCES

McFADDEN, J. A. (1956). An approximation for the symmetric, quadrivariate normal integral. *Biometrika*, **43**, 206–7.

MORAN, P. A. P. (1956). The numerical evaluation of a class of integrals. *Proc. Camb. Phil. Soc.* **52**, 230–3.

PRICE, R. (1959). A useful theorem for nonlinear devices having Gaussian inputs. *I.R.E. Transactions on Information Theory*, **4**, 69–73.

SHANKS, D. (1955). Nonlinear transformations of divergent and slowly convergent sequences. *J. Math. Phys.* **34**, 1–42.

On the stochastic matrix in a genetic model of Moran

BY J. GANI†

1. INTRODUCTION

In some recent work Moran (1958) has used a Markov chain embedded in a birth and death process to construct a genetic model for the random fluctuation of gene frequencies in a population of haploid monoecious individuals subject to mutation and selection. The model consists of a fixed number M of gametes of type a or A, and the system is said to be in state j when j of these are a, and $M-j$ are A. If the probabilities of mutation of a gamete a to A and A to a are respectively α_1, α_2 $(0 \leqslant \alpha_1, \alpha_2 \leqslant 1)$, the transition probabilities p_{jk} from state j to state k of the stochastic matrix

$$P(\alpha_1, \alpha_2) = \{p_{jk}(\alpha_1, \alpha_2)\} \quad (j, k = 0, 1, \ldots, M)$$

for the embedded Markov chain of the system are shown to be

$$p_{j,j-1} = \frac{j}{M} q_j, \quad p_{j,j} = \frac{j}{M} p_j + \left(1 - \frac{j}{M}\right) q_j, \quad p_{j,j+1} = \left(1 - \frac{j}{M}\right) p_j, \quad p_{jk} = 0 \quad (k < j-1,\ k > j+1), \tag{1·1}$$

where

$$p_j = \frac{j}{M}(1-\alpha_1) + \left(1 - \frac{j}{M}\right)\alpha_2, \quad q_j = \frac{j}{M}\alpha_1 + \left(1 - \frac{j}{M}\right)(1-\alpha_2). \tag{1·2}$$

The essential problem is to determine the rate of progress of the system to homozygosity; this depends on the largest characteristic root smaller than $\lambda_0(\alpha_1, \alpha_2) = 1$ of the matrix $P(\alpha_1, \alpha_2)$. When $\alpha_1 = \alpha_2 = 0$, Hannan (see Appendix to Moran, 1958) has shown that the roots of the matrix $P(0, 0)$ are

$$\lambda_j(0, 0) = 1 - \frac{j(j-1)}{M^2} \quad (j = 0, 1, 2, \ldots, M) \tag{1·3}$$

so that in this case progress to homozygosity will depend on the largest non-unit root

$$\lambda_2(0, 0) = 1 - \frac{2}{M^2}.$$

For $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, Moran uses an interesting 'expectation' technique to obtain information about the roots of the stochastic matrix: he considers the random variable $X_t = jM^{-1}$ after the tth transition, and proves that

$$\left.\begin{aligned} E(X_{t+1}|X_t) &= \alpha_2 M^{-1} + X_t[1 - M^{-1}(\alpha_1 + \alpha_2)], \\ E(X_{t+1}^2|X_t) &= \alpha_2 M^{-2} + X_t[2\alpha_2 M^{-1} + M^{-2}(2 - \alpha_1 - 3\alpha_2)] \\ &\quad + X_t^2[1 - 2M^{-1}(\alpha_1 + \alpha_2) + 2M^{-2}(1 - \alpha_1 - \alpha_2)]. \end{aligned}\right\} \tag{1·4}$$

From these results, he surmises that the coefficients of X_t in $E(X_{t+1}|X_t)$ and X_t^2 in $E(X_{t+1}^2|X_t)$ are respectively

$$\left.\begin{aligned} \lambda_1(\alpha_1, \alpha_2) &= 1 - M^{-1}(\alpha_1 + \alpha_2), \\ \lambda_2(\alpha_1, \alpha_2) &= 1 - 2M^{-1}(\alpha_1 + \alpha_2) + 2M^{-2}(1 - \alpha_1 - \alpha_2), \end{aligned}\right\} \tag{1·5}$$

and concludes that since their values agree for $\alpha_1 = \alpha_2 = 0$ with Hannan's results, his conjecture will be true at least for small α_1, α_2.

† Presently at the Australian National University; this work was partly done while the author was visiting Stanford University on an NSF grant.

We show in this paper that for $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ the latent roots $\lambda_j(\alpha_1, \alpha_2)$ must all be real and distinct, so that in fact Moran's conjecture is exact. We also prove that if, as for the present model, we can express $E(X_{t+1}^k|X_t)$ as

$$E(X_{t+1}^k|X_t) = a_{0k} + a_{1k}X_t + \ldots + a_{kk}X_t^k \quad (k = 0, 1, \ldots, M), \tag{1·6}$$

where the $a_{jk} = a_{jk}(\alpha_1, \alpha_2)$, then it follows that the a_{kk} are the required latent roots of the matrix.

2. Properties of the roots of $P(\alpha_1, \alpha_2)$

It is well known that the stochastic matrix P for a positively regular Markov chain with $M+1$ states has latent roots

$$\lambda_0 = 1, \quad 1 > |\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_M|, \tag{2·1}$$

where the λ_j $(j = 2, \ldots, M)$ may be real or complex.

We shall prove that if, for a positively regular Markov chain, P is the continuant matrix of transition probabilities

$$P = \begin{bmatrix} p_{00} & p_{01} & 0 & \ldots & \ldots \\ p_{10} & p_{11} & p_{12} & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & 0 & p_{M-1,M-2} & p_{M-1,M-1} & p_{M-1,M} \\ \ldots & \ldots & 0 & p_{M,M-1} & p_{MM} \end{bmatrix} \tag{2·2}$$

with non-zero off-diagonal elements $p_{j,j-1}$ $(j = 1, 2, \ldots, M)$, $p_{j,j+1}$ $(j = 0, 1, \ldots, M-1)$, then the roots of the characteristic equation

$$A_{M+1}(\lambda) = |\lambda I - P| = 0 \tag{2·3}$$

are real and distinct, so that

$$\lambda_0 = 1 > \lambda_1 > \lambda_2 > \ldots > \lambda_M > -1. \tag{2·4}$$

As the proof is similar to that given by Ledermann & Reuter (1954) for the roots of the matrix associated with a birth and death process, it will be only briefly outlined here.

Let $A_n(\lambda) = |\lambda I^{(n)} - P^{(n)}|$ $(n = 1, 2, \ldots, M)$ be the nth leading minor of the determinant $A_{M+1}(\lambda)$ starting from the top. Then, defining $A_0(\lambda) = 1$, we have that

$$\left.\begin{aligned} A_1(\lambda) &= \lambda - p_{00}, \\ A_2(\lambda) &= (\lambda - p_{00})(\lambda - p_{11}) - p_{10}p_{01} \end{aligned}\right\} \tag{2·5}$$

and so on, where $A_n(\lambda)$ will clearly be a polynomial of degree n in λ. It may readily be shown by expanding the minor $A_n(\lambda)$ $(n = 2, 3, \ldots, M)$ that

$$A_n(\lambda) = (\lambda - p_{n-1,n-1})A_{n-1}(\lambda) - p_{n-1,n-2}p_{n-2,n-1}A_{n-2}(\lambda), \tag{2·6}$$

where $p_{n-1,n-2}p_{n-2,n-1} > 0$.

We illustrate the general argument by applying it to $A_2(\lambda)$; from (2·6) we may write

$$A_2(\lambda) = (\lambda - p_{11})A_1(\lambda) - p_{10}p_{01}A_0. \tag{2·7}$$

Now $A_1(\lambda)$ has one real root p_{00}, and when $\lambda = p_{00}$ we have that

$$A_2(p_{00}) = -p_{10}p_{01}A_0 < 0,$$

since $A_0 > 0$. As $A_2(\lambda)$ is a quadratic in λ, and $A(\pm\infty) = \infty$, it follows that $A_2(\lambda)$ has two real roots separated by p_{00}, the root of $A_1(\lambda)$.

In general, if the polynomials $A_{2r-1}(\lambda)$, $A_{2r}(\lambda)$ $(r = 1, 2, \ldots, [\frac{1}{2}M])$ exhibit the sign patterns in (2·8), where the real roots of $A_{2r-1}(\lambda)$ separate those of $A_{2r}(\lambda)$, then from (2·6) it follows that $A_{2r+1}(\lambda)$ will also have the sign pattern given.

$$\left.\begin{aligned} A_{2r-1}(\lambda)&: \quad - \; 0 \; + \; 0 \; - \quad \ldots \quad + \; 0 \; - \; 0 \; + \\ A_{2r}(\lambda)&: \quad + \; 0 \; - \; 0 \; + \; 0 \; - \ldots - \; 0 \; + \; 0 \; - \; 0 \; + \\ \hline A_{2r+1}(\lambda)&: \quad - \; 0 \; + \; 0 \; - \; 0 \; + \quad \ldots \quad - \; 0 \; + \; 0 \; - \; 0 \; + \end{aligned}\right\} \tag{2·8}$$

For, at the zeros of $A_{2r}(\lambda)$, $A_{2r+1}(\lambda)$ will always have a sign opposite to that of $A_{2r-1}(\lambda)$, and since $A_{2r+1}(\infty) = \infty$, $A_{2r+1}(-\infty) = -\infty$, there must be two further zeros in the two end regions where $A_{2r}(\lambda) > 0$. A similar argument may now be extended to $A_{2r+2}(\lambda)$. By induction it follows that $A_{M+1}(\lambda)$ will have $M+1$ real and distinct roots. We know further that $\lambda_0 = 1$ and $|\lambda_j| < 1$ $(j = 1, \ldots, M)$, so that the statement (2·4) is proved.

In Moran's model, the probabilities p_{jk} are of the form (1·1), with off-diagonal elements of the stochastic matrix $P(\alpha_1, \alpha_2)$ being non-zero when $\alpha_1 \neq 0$, $\alpha_2 \neq 0$; hence the roots $\lambda_j(\alpha_1, \alpha_2)$ $(j = 0, 1, \ldots, M)$ of this continuant matrix will be real and distinct.

When $\alpha_1 = \alpha_2 = 0$, $p_{01} = p_{M,M-1} = 0$, and the roots of the matrix $P(0, 0)$ as obtained by Hannan are

$$\lambda_0(0, 0) = \lambda_1(0, 0) = 1, \quad \lambda_j(0, 0) = 1 - \frac{j(j-1)}{M^2} \quad (j = 2, \ldots, M). \tag{2·9}$$

Using his expectation technique, Moran finds two roots

$$\left.\begin{aligned} a_{11}(\alpha_1,\alpha_2) &= 1-M^{-1}(\alpha_1+\alpha_2),\\ a_{22}(\alpha_1,\alpha_2) &= 1-2M^{-1}(\alpha_1+\alpha_2)+2M^{-2}(1-\alpha_1-\alpha_2)\end{aligned}\right\} \tag{2·10}$$

of the matrix $P(\alpha_1,\alpha_2)$, and conjectures that these are respectively $\lambda_1(\alpha_1,\alpha_2)$ and $\lambda_2(\alpha_1,\alpha_2)$, at least for small α_1, α_2.

Since the roots of $P(\alpha_1,\alpha_2)$ are real, distinct and continuous in α_1, α_2 and since

$$a_{11}(0,0)=1=\lambda_1(0,0),\quad a_{22}(0,0)=1-2M^{-2}=\lambda_2(0,0) \tag{2·11}$$

it will follow that for all values of $\alpha_1,\alpha_2>0$

$$a_{11}(\alpha_1,\alpha_2)=\lambda_1(\alpha_1,\alpha_2),\quad a_{22}(\alpha_1,\alpha_2)=\lambda_2(\alpha_1,\alpha_2). \tag{2·12}$$

Thus Moran's conjecture is exact.

3. The expectation technique

Moran's expectation technique depends on the fact that, for the particular continuant matrix $P(\alpha_1,\alpha_2)$ with elements (1·1), it is possible (cf. equations (1·4)) to express $E(X_{t+1}|X_t)$ and $E(X_{t+1}^2|X_t)$ respectively as linear and quadratic functions of X_t

$$\left.\begin{aligned} E(X_{t+1}|X_t) &= a_{01}+a_{11}X_t,\\ E(X_{t+1}^2|X_t) &= a_{02}+a_{12}X_t+a_{22}X_t^2.\end{aligned}\right\} \tag{3·1}$$

Here, the $a_{ij}=a_{ij}(\alpha_1,\alpha_2)$ are functions of α_1, α_2, and $X_t=jM^{-1}$ $(j=0,1,\ldots,M)$ is the random variable giving the proportion of gametes a in a population of size M after t transitions of the system.

We note from (1·1) and (1·2) that the transition probabilities are themselves quadratic functions of X_t

$$\left.\begin{aligned} p_{X_t,X_t-M^{-1}} &= (1-\alpha_2)X_t-(1-\alpha_1-\alpha_2)X_t^2 && (X_t=M^{-1},2M^{-1},\ldots,1),\\ p_{X_t,X_t} &= (1-\alpha_2)+(\alpha_1+3\alpha_2-2)X_t+2(1-\alpha_1-\alpha_2)X_t^2 && (X_t=0,M^{-1},\ldots,1),\\ p_{X_t,X_t+M^{-1}} &= \alpha_2+(1-\alpha_1-2\alpha_2)X_t-(1-\alpha_1-\alpha_2)X_t^2 && (X_t=0,M^{-1},\ldots,(M-1)M^{-1}).\end{aligned}\right\} \tag{3·2}$$

Their values are such that

$$\left.\begin{aligned} E(X_{t+1}-X_t|X_t) &= M^{-1}(p_{X_t,X_t+M^{-1}}-p_{X_t,X_t-M^{-1}})\\ &= M^{-1}[\alpha_2-(\alpha_1+\alpha_2)X_t],\\ E[(X_{t+1}-X_t)^2|X_t] &= M^{-2}(p_{X_t,X_t+M^{-1}}+p_{X_t,X_t-M^{-1}})\\ &= M^{-2}[\alpha_2+(2-\alpha_1-3\alpha_2)X_t-2(1-\alpha_1-\alpha_2)X_t^2].\end{aligned}\right\} \tag{3·3}$$

From these, one may derive equations (1·4) which are clearly of the form (3·1).

Since $X_{t+1}-X_t$ can only take the values $-M^{-1}$, 0, M^{-1}, then for all integral $r\geqslant 2$,

$$\begin{aligned} E[(X_{t+1}-X_t)^{2r-1}|X_t] &= M^{-(2r-1)}[\alpha_1-(\alpha_1+\alpha_2)X_t],\\ E[(X_{t+1}-X_t)^{2r}|X_t] &= M^{-2r}[\alpha_2+(2-\alpha_1-3\alpha_2)X_t-2(1-\alpha_1-\alpha_2)X_t^2].\end{aligned}$$

Hence

$$\left.\begin{aligned} E(X_{t+1}^{2r-1}|X_t) &= \sum_{j=1}^{2r-1}\binom{2r-1}{j}(-1)^{j-1}X_t^jE(X_{t+1}^{2r-1-j})+M^{-(2r-1)}[\alpha_1-(\alpha_1+\alpha_2)X_t],\\ E(X_{t+1}^{2r}|X_t) &= \sum_{j=1}^{2r}\binom{2r}{j}(-1)^{j-1}X_t^jE(X_{t+1}^{2r-j})+M^{-2r}[\alpha_2+(2-\alpha_1-3\alpha_2)X_t-2(1-\alpha_1-\alpha_2)X_t^2].\end{aligned}\right\} \tag{3·4}$$

Thus it is readily seen that for $k=0,1,2,\ldots,M$, we obtain

$$E(X_{t+1}^k|X_t)=a_{0k}+a_{1k}X_t+\ldots+a_{kk}X_t^k, \tag{3·5}$$

where the $a_{jk}=a_{jk}(\alpha_1,\alpha_2)$.

These relations may be expressed in the matrix form

$$P(\alpha_1,\alpha_2)X=XA, \tag{3·6}$$

where the matrices

$$X=\begin{bmatrix} 1 & 0 & 0 & \ldots & 0\\ 1 & M^{-1} & M^{-2} & \ldots & M^{-M}\\ 1 & 2M^{-1} & (2M^{-1})^2 & \ldots & (2M^{-1})^M\\ \vdots & & & & \\ 1 & (M-1)M^{-1} & [(M-1)M^{-1}]^2 & \ldots & [(M-1)M^{-1}]^M\\ 1 & 1 & 1 & \ldots & 1\end{bmatrix},\quad A=\begin{bmatrix} a_{00} & a_{01} & a_{02} & \ldots & a_{0M}\\ 0 & a_{11} & a_{12} & \ldots & a_{1M}\\ \ldots & 0 & a_{22} & \ldots & a_{2M}\\ & & & & \\ \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & 0 & \ldots & a_{MM}\end{bmatrix}$$

have elements $X_{jk} = (jM^{-1})^k$, $A_{jk} = a_{jk}(\alpha_1, \alpha_2)$ $(j, k = 0, 1, \ldots, M)$, respectively, A being upper triangular with $a_{00} = 1$. The matrix X clearly consists of $M+1$ linearly independent vectors, and will be non-singular.

We may therefore write

$$P(\alpha_1, \alpha_2) = XAX^{-1} \tag{3·7}$$

and it follows directly that the latent roots $\lambda_j(\alpha_1, \alpha_2)$ will also be those of A, namely the diagonal elements $a_{kk}(\alpha_1, \alpha_2)$.

4. Explicit values for the roots of $P(\alpha_1, \alpha_2)$

We now use the expectation technique to obtain all roots of $P(\alpha_1, \alpha_2)$. From (3·4), we see that for $k = 3, 4, \ldots, M$, the coefficient of X_t^k in $E(X_{t+1}^k | X_t)$ will be

$$a_{kk} = \sum_{j=1}^{k} \binom{k}{j} (-1)^{j-1} a_{k-j\ k-j}, \tag{4·1}$$

and it is readily proved by induction that

$$a_{kk}(\alpha_1, \alpha_2) = 1 - kM^{-1}(\alpha_1 + \alpha_2) - k(k-1)\,M^{-2}(1 - \alpha_1 - \alpha_2). \tag{4·2}$$

For if the diagonal elements a_{jj} of A are of the form (4·2) for $j = 0, 1, 2, \ldots, k-1 < M$ then

$$a_{kk} = \sum_{j=1}^{k} \binom{k}{j} (-1)^{j-1} \{1 - (k-j)\,M^{-1}(\alpha_1 + \alpha_2) - (k-j)\,(k-j-1)\,M^{-2}(1 - \alpha_1 - \alpha_2)\}. \tag{4·3}$$

This is readily reducible to (4·2) by using the identity

$$(x-1)^k = x^k + \sum_{j=1}^{k} \binom{k}{j} (-1)^j x^{k-j}$$

and its first and second derivatives with respect to x, when $x = 1$.

From § 3, the a_{kk} of (4·2) are in fact known to be the latent roots $\lambda_j(\alpha_1, \alpha_2)$ of $P(\alpha_1, \alpha_2)$. It should be noted that in this particular case $a_{k+1,\,k+1} < a_{kk}$ for all $k = 0, 1, \ldots, M-1$ since

$$\begin{aligned} a_{kk} - a_{k+1,\,k+1} &= M^{-1}(\alpha_1 + \alpha_2) + 2kM^{-2}(1 - \alpha_1 - \alpha_2) \\ &= (\alpha_1 + \alpha_2)\,M^{-1}(1 - kM^{-1}) + kM^{-2}(2 - \alpha_1 - \alpha_2) > 0 \end{aligned} \tag{4·4}$$

so that the a_{kk} are also ordered, and so are precisely the roots

$$\lambda_k(\alpha_1, \alpha_2) = a_{kk}(\alpha_1, \alpha_2) \quad (k = 0, 1, \ldots, M).$$

For $\alpha_1 = \alpha_2 = 0$, these reduce to Hannan's roots (2·9).

Similar results for the stochastic matrices of Moran's more complex models, in which the states of Markov chains are defined by two and four variates respectively, will be given in a further paper.

References

Moran, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60–71.

Ledermann, W. & Reuter, G. E. H. (1954). Spectral theory for the differential equations of simple birth and death processes. *Phil. Trans.* A, **246**, 321–69.

Note added in proof: For a comprehensive analysis of Moran's model in continuous time, the reader is referred to:

Karlin, S. & McGregor, J. (1960). On a genetics model of Moran. Stanford University Technical Report.

Departures from assumption in sequential analysis

By W. J. Ewens

Department of Statistics, University of Melbourne

I. Introduction

In carrying out a sequential test, one is normally testing for the value of a parameter in a distribution, the density type involved being assumed known. This paper is concerned with the effect on such a test of wrongly specifying this density type. Particular reference is made to the case where the assumption of normality is made. We distinguish two cases; testing for the mean of a population, and testing for

the variance. A single formula is derived, covering both cases, from which complete sets of power and average sample size curves may be found. While the examples given cover only these two cases, the formula is valid for departures from any assumed distribution.

As in the case of fixed-sample-size tests, we find that the sequential test for means is comparatively robust in respect to departure from assumed normality but the test for variances is very sensitive.

2. Sequential testing

If we wish to test sequentially the hypothesis

$$H_0\colon f(x) \equiv f(x, \theta_0),$$

against

$$H_1\colon f(x) \equiv f(x, \theta_1),$$

with strength (α, β), we continue testing while

$$\beta/(1-\alpha) < \prod_{i=1}^{n} f(x_i, \theta_1)/f(x_i, \theta_0) < (1-\beta)/\alpha \tag{1}$$

accepting $H_0(H_1)$ if the lower (upper) inequality is broken. Equation (1) is more conveniently written in terms of logarithms

$$\ln\{\beta/(1-\alpha)\} < Z_n = \sum_{i=1}^{n} z_i < \ln\{(1-\beta)/\alpha\}, \tag{2}$$

where

$$z_i = \ln\{f(x_i, \theta_1)/f(x_i, \theta_0)\}.$$

Wald (1947) has shown that if the true value of the parameter is θ, then P = Prob. (accepting H_0) is given by

$$P = \frac{\{(1-\beta)/\alpha\}^h - 1}{\{(1-\beta)/\alpha\}^h - \{\beta/(1-\alpha)\}^h}, \tag{3}$$

where h is the unique non-zero solution for t in

$$\int_{-\infty}^{+\infty} \{f(x, \theta_1)/f(x, \theta_0)\}^t f(x, \theta)\, dx = 1. \tag{4}$$

Such a value h can be shown to exist under fairly general conditions. The average sample number (A.S.N.) when the parameter takes the value θ is

$$\text{A.S.N.} = \frac{P \ln\{\beta/(1-\alpha)\} + (1-P)\ln\{(1-\beta)/\alpha\}}{E(z|\theta)}, \tag{5}$$

where P is found from (3) and (4). In the case $E(z|\theta) = 0$,

$$P = \frac{\ln\{(1-\beta)/\alpha\}}{\ln\{(1-\beta)/\alpha\} - \ln\{\beta/(1-\alpha)\}},$$

$$\text{A.S.N.} = \frac{-\ln\{(1-\beta)/\alpha\}\ln\{\beta/(1-\alpha)\}}{E(z^2|\theta)}.$$

It may be shown that if the null and alternative hypotheses are

$$H_0\colon x \cap N(\theta_0, \sigma^2),$$

$$H_1\colon x \cap N(\theta_1, \sigma^2),$$

then

$$h = (\theta_0 + \theta_1 - 2\theta)/(\theta_1 - \theta_0),$$

$$E(z|\theta) = (\theta_1 - \theta_0)(2\theta - \theta_0 - \theta_1)/(2\sigma^2).$$

Similarly, if null and alternative hypotheses are

$$H_0\colon x \cap N(0, \sigma_0^2),$$

$$H_1\colon x \cap N(0, \sigma_1^2)$$

equation (4) reduces to

$$(\sigma_0^2/\sigma_1^2)^t = 1 - t\sigma^2(1/\sigma_0^2 - 1/\sigma_1^2),$$

where σ^2 is the true variance value. The A.S.N. is given by (5) if we put

$$E(z|\sigma^2) = \tfrac{1}{2}\sigma^2(1/\sigma_0^2 - 1/\sigma_1^2) - \ln\{\sigma_1/\sigma_0\}.$$

3. Departures from assumed distributions

We now aim to generalize the above formulae to the case where the density type is wrongly specified. In doing we so will use neither Wald's fundamental identity nor his artificial density functions, but simple difference equations.

We suppose that the null and alternative hypotheses are

$$H_0\colon f(x) \equiv f(x, \theta_0),$$
$$H_1\colon f(x) \equiv f(x, \theta_1)$$

and that the test is of strength (α, β). Suppose that x has the density $g(x, \theta)$, which may or may not be of the same type as f. Then $z = \ln\{f(x, \theta_1)/f(x, \theta_0)\}$ will have density $\phi(z)$ given by

$$g(x)\,dx = \phi(z)\,dz, \quad z = z(x).$$

The test is naturally carried out in terms of Σz_i, since we assume the densities f in null and alternative hypothesis. We let $p(y)$ denote the probability of accepting H_0 when the cumulative sum Σz_i is y, and $n(y)$ the expected further sample size at this point. Then if we make the usual assumptions about negligible boundary overlap, we have

$$p(y) = \int_{-\infty}^{+\infty} p(y+z)\,\phi(z)\,dz, \tag{5}$$

$$n(y) = \int_{-\infty}^{+\infty} n(y+z)\,\phi(z)\,dz + 1. \tag{6}$$

Putting $p(y) = \exp(ty)$, we have

$$1 = \int_{-\infty}^{+\infty} \exp(tz)\,\phi(z)\,dz. \tag{7}$$

It may be shown that under similar conditions to those of Wald (1947), (7) has a unique non-zero solution for t. If we denote this solution by h, we may find the power curve by applying to the general solution $p(y) = A + B\exp(hy)$ the boundary conditions

$$p\{\ln\beta/(1-\alpha)\} = 1, \quad p\{\ln(1-\beta)/\alpha\} = 0.$$

These conditions fix the constants A and B, giving

$$\begin{aligned} P \equiv \text{Prob.(accepting } H_0) &= p(0) \\ &= \frac{\exp[h\ln\{(1-\beta)/\alpha\}] - 1}{\exp[h\ln\{(1-\beta)/\alpha\}] - \exp[h\ln\{\beta/(1-\alpha)\}]} \\ &= \frac{\{(1-\beta)/\alpha\}^h - 1}{\{(1-\beta)/\alpha\}^h - \{\beta/(1-\alpha)\}^h}. \end{aligned} \tag{8}$$

Since (7) may be rewritten

$$1 = \int_{-\infty}^{+\infty} \{f(x, \theta_1)/f(x, \theta_0)\}^t\, g(x, \theta)\,dx \tag{9}$$

we find that (8) and (9) are sufficient to determine the power curve. We need only put any value of θ required in (9), solve for t, and substitute in (8). This generalizes the usual power curve formula. The A.S.N. curve is found as follows. Equation (6) is non-homogeneous, with particular solution

$$n(y) = -y/E(z).$$

General solutions are therefore $n(y) = -y/E(z) + C + D\exp(hy)$,

where C and D are arbitrary constants, and h is given by (7) or (9). C and D are fixed by the boundary conditions

$$n[\ln\{\beta/(1-\alpha)\}] = n[\ln\{(1-\beta)/\alpha\}] = 0,$$

giving

$$E(n) = n(0) = \frac{P\ln\{\beta/(1-\alpha)\} + (1-P)\ln\{(1-\beta)/\alpha\}}{E(z)}, \tag{10}$$

where P is found from (8) and (9). Thus the usual formula for the A.S.N. continues to hold if P and $E(z)$ are interpreted correctly.

The most interesting applications of the above formulae are made when both null and alternative hypotheses assume a normal distribution of the observations, since in this case we may compare the results with well-known fixed-sample-size test results. We consider both tests for means and tests for variances.

(*a*) *Tests for means*

(i) We first suppose that the true density of the variate is the double-exponential density

$$g(x) = \exp[-\sqrt{2}\,|x-\mu|/\sigma]/(\sqrt{2}\,\sigma) \quad (-\infty < x < +\infty),$$

for which $E(x) = \mu$, $\text{var}(x) = \sigma^2$, $\gamma_1 = 0$, $\gamma_2 = 3$. Suppose also that null and alternative hypotheses are

$$H_0\colon x \cap N(-\tfrac{1}{2}, \sigma^2),$$
$$H_1\colon x \cap N(+\tfrac{1}{2}, \sigma^2).$$

After some reduction (9) becomes

$$\exp(\mu t/\sigma^2) = 1 - t^2/(2\sigma^2). \tag{11}$$

On the other hand, if x were normally distributed with mean μ and variance σ^2, equation (9) becomes

$$-2\mu = t. \tag{12}$$

Let us fix σ^2 at any arbitrary value. Then for any value of μ, we may solve both (11) and (12) for t, and hence find two values of P. By varying μ we generate two distinct power curves. Finally, we may change σ^2 and generate a further set of two curves. The results are tabled below for $\alpha = \beta = 0{\cdot}05$. It is clear from (11) and (12) that $\mu = -\mu'$ gives $t = -t'$ and hence $P = 1 - P'$. Therefore only half of each curve need be given.

Table 1. *Probability of rejecting the null hypothesis* $\{x \cap N(-\frac{1}{2}, \sigma^2)\}$ *in favour of the alternative* $\{x \cap N(+\frac{1}{2}, \sigma^2)\}$, *the population sampled being* (*a*) *double-exponential* (D.E.), (*b*) *normal, with indicated mean and variance.*

$\alpha = \beta = 0{\cdot}05$.

$\sigma^2 = 4$			$\sigma^2 = 1$		
Mean	D.E.	Normal	Mean	D.E.	Normal
0·662	0·972	0·980	1·061	0·972	0·998
·594	·962	·971	0·844	·962	·993
·534	·950	·959	·693	·950	·983
·474	·934	·942	·577	·934	·968
·417	·913	·921	·482	·913	·945
0·358	0·887	0·892	0·402	0·887	0·914
·307	·854	·859	·331	·854	·875
·254	·813	·817	·267	·813	·828
·202	·765	·767	·209	·765	·773
·151	·708	·708	·154	·708	·712
0·100	0·643	0·643	0·101	0·643	0·644
·050	·573	·573	·050	·573	·573
·000	·500	·500	·000	·500	·500

Table 2. *Probability of rejecting the null hypothesis* $\{x \cap N(-\frac{1}{2}, 1)\}$ *in favour of the alternative* $\{x \cap N(+\frac{1}{2}, 1)\}$, *the population sampled being* (*a*) *rectangular*, (*b*) *normal, with variance* 1 *and indicated mean.*

$\alpha = \beta = 0{\cdot}05$.

Mean	Rectangular	Normal
0·530	0·972	0·958
·496	·962	·949
·457	·950	·936
·417	·934	·921
·376	·913	·901
0·334	0·887	0·878
·290	·854	·846
·244	·813	·808
·197	·765	·761
·149	·708	·706
0·099	0·643	0·642
·050	·573	·573
·000	·500	·500

(ii) We now suppose that the true density of the observations is of the rectangular form

$$g(x) = 1/(2\sqrt{3}\sigma), \quad (\mu - \sqrt{3}\,\sigma \leqslant x \leqslant \mu + \sqrt{3}\,\sigma).$$

Let H_0, H_1, α and β be as in the previous example. Then (9) becomes

$$\exp(\mu t/\sigma^2) = \sqrt{3}\,t/\{\sigma \sinh(\sqrt{3}\,t/\sigma)\}. \tag{13}$$

As before, we first fix σ, and then solve (13) for various μ. It is also evident from (13) that the symmetry conditions of the previous example also apply here.

We observe that if we base the sequential scheme on the usual assumption of normality, the probability levels for both populations are not far from those which hold in the case where the distribution of the observations is, in fact, normal.† Since one of the populations considered has high positive kurtosis (+3) and the other negative kurtosis (−1·2) we may infer that population kurtosis does not markedly affect the test. This result agrees with the corresponding conclusion for fixed-sample-size tests (cf. Pearson, 1931; Geary, 1947; Gayen, 1950; Box & Andersen, 1955).

(*b*) *Tests for variances*

For tests on variances, we have

$$\left.\begin{aligned} H_0&: x \cap N(0, \sigma_0^2), \\ H_1&: x \cap N(0, \sigma_1^2). \end{aligned}\right\} \tag{14}$$

We have shown that if $x \cap N(0, \sigma^2)$ and the test is of strength (α, β), then P is given by equation (8), where h is the non-zero solution for t in

$$(\sigma_0^2/\sigma_1^2)^t = 1 - t\sigma^2(1/\sigma_0^2 - 1/\sigma_1^2). \tag{15}$$

We can now use (9) to extend this to the case of non-normal observations.

(i) Let us consider the double-χ density

$$g(x) = \left\{\frac{\delta}{2\sigma^2}\right\}^{\frac{1}{2}\delta} \frac{1}{\Gamma(\frac{1}{2}\delta)} |x|^{\delta-1} \exp(-\tfrac{1}{2}x^2\delta/\sigma^2), \tag{16}$$

$$E(x^{2r+1}) = 0, \quad E(x^2) = \sigma^2, \quad \delta^{-1} = 1 + \tfrac{1}{2}\gamma_2.$$

Table 3. *Probability of rejecting the null hypothesis, for various variance values, the variate being from a double-χ density with the indicated value of kurtosis.*

$\alpha = \beta = 0{\cdot}05, \quad \sigma_0^2 = 1, \quad \sigma_1^2 = 4.$

σ^2 \ γ_2	−1	0 (normal)	1	2
4·752	0·99915	0·972	0·913	0·854
4·000	·9972	·950	·877	·813
3·386	·991	·913	·828	·765
2·883	·972	·854	·765	·708
2·476	·913	·765	·687	·643
2·130	0·765	0·643	0·597	0·573
1·848	·500	·500	·500	·500
1·616	·235	·357	·403	·427
1·421	·087	·235	·313	·357
1·255	·028	·146	·235	·292
1·117	0·0089	0·087	0·172	0·235
1·000	·00276	·050	·123	·187
0·905	·00085	·028	·087	·146

If we have null and alternative hypotheses as specified by (14), and use (16) for $g(x)$ in (9), then after a little reduction the following equation for t is found

$$(\sigma_0^2/\sigma_1^2)^{t/\delta} = 1 - (1/\sigma_0^2 - 1/\sigma_1^2)\, t\sigma^2/\delta \tag{17}$$

which is clearly a generalization of (15). Equations (3) and (17) may be rewritten

$$P = \frac{\{(1-\beta)/\alpha\}^{h\delta} - 1}{\{(1-\beta)/\alpha\}^{h\delta} - \{\beta/(1-\alpha)\}^{h\delta}}, \tag{18}$$

where h is the non-zero solution of (15) for t. Since $\delta = (1 + \frac{1}{2}\gamma_2)^{-1}$, we find a remarkable analogy between this result and the corresponding one for fixed-sample-size tests. Here Box (1953) has shown that the test statistic M_1 (used in the test for heterogeneity of variance) computed from a double-χ distribution is distributed as $(1 + \frac{1}{2}\gamma_2)M_1$ computed from a normal population. One may also conjecture that (18) is 'asymptotically' true for any distribution with finite cumulants. ('Asymptotic' here meaning for

† Agreement is not so good in the double exponential case when $\sigma = 1$ as when $\sigma = 4$, but the discrepancy is far less than that obtained below when testing for variance.

($\sigma_0^2 - \sigma_1^2$) small.) Equation (18) may now be used to compute power curves for various values of γ_2. The results are given in Table 3. Evidently non-zero values of kurtosis upset the probability levels to a marked degree. This result agrees with the corresponding behaviour of the test statistic for fixed-sample-size tests (cf. Box, (1953). In this paper the sequential test power curve is evaluated at the points where H_0 and H_1 are true, but by a different method than above.)

(c) *Average sample size*

We may use the above power curve tables, in conjunction with equation (10), to find the A.S.N. curves when normality is wrongly assumed in sequential tests.

Since tests on means show small divergence in power curves from the nominal, the A.S.N. curves will also show only small divergence. However, there will naturally be large divergences for the test on variances, since in this case the power curves differ markedly. We use Table 3 and equation (10) to derive Table 4 below, which gives the A.S.N. curves for the distributions involved in Table 3. An interesting result is that for the case $\gamma_2 = 2$, the maximum A.S.N. occurs outside the range (σ_0^2, σ_1^2). A joint examination of Tables 3 and 4 shows the extent to which 'normal theory' has now broken down.

Table 4. *Average sample sizes corresponding to the distributions of Table* 3

σ^2 \ γ_2	−1	0 (normal)	1	2
4·752	2·70	2·55	2·23	1·91
4·000	3·63	3·28	2·75	2·28
3·386	5·01	4·22	3·35	2·71
2·883	7·16	5·37	4·02	3·16
2·476	10·33	6·63	4·68	3·58
2·130	14·76	7·96	5·40	4·07
1·848	18·06	9·03	6·02	4·51
1·615	17·86	9·64	6·54	4·92
1·421	15·19	9·75	6·88	5·26
1·255	12·50	9·37	7·02	5·51
1·117	10·55	8·87	7·05	5·69
1·000	9·21	8·33	6·98	5·81
0·905	8·89	7·86	6·88	5·90

5. Summary

Comparison has been made between the effects of non-normality in a fixed-sample-size test and in a sequential test. Although different methods of investigation are necessary, the results obtained are similar. In both tests there is a comparative insensitivity to non-normality for tests on means, and a sensitivity to non-zero values of kurtosis for tests on variances. The tests have a similar asymptotic property. A formula for the A.S.N. curve for the sequential test has been derived.

References

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, **40**, 318–35.

Box, G. E. P. & Andersen, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *J. R. Statist. Soc.* B, **17**, 1–26.

Gayen, A. K. (1950). Significance of the difference between the means of two non-normal samples. *Biometrika*, **37**, 399–408.

Geary, R. C. (1947). Testing for normality. *Biometrika*, **34**, 209–42.

Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, **23**, 114–33.

Wald, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons, Inc.

A note on some asymptotic properties of the logarithmic series distribution

By J. C. GOWER

Rothamsted Experimental Station

Summary

Formulae are given to calculate

$$S(x, R) = \sum_{r=R+1}^{\infty} \frac{x^r}{r},$$

the remainder term in the logarithmic series, which are valid for high values of R and values of x close to unity. These are of value when dealing with extreme values of the parameters appearing in the logarithmic series distribution such as have arisen in C. B. Williams's recent study of the possible distribution of the number per species for all the insects in the world. Bounds are given for the errors involved when making the recommended approximations. The calculation of the median of the distribution is also discussed.

Introduction

The Logarithmic Series Distribution was introduced in an article by Fisher, Corbett & Williams (1943) who fitted it to the numbers of individuals per species in samples of Macro-Lepidoptera caught in a light trap at Rothamsted. Since then a number of papers, notably those of C. B. Williams, have used this distribution to describe data of diverse kinds. Recently Williams (1960) has investigated the problem of assessing the possible distribution of the number per species of all the insects in the world. Attention is focused on the number of species with more than a specified number of individuals. When this specified number is very large and the logarithmic distribution is used, a knowledge of its asymptotic properties is needed. Williams (1960) remarks that 'this is a complex problem which requires further investigation'. In this article some asymptotic properties of the logarithmic distribution will be listed and some indication is given of the range of values of the parameters of this distribution for which the results are reasonably accurate. Biologists should note these restrictions, since if the formulae are applied outside the range for which they are designed, inaccurate results may be obtained.

Using the standard notation, suppose we have N individuals distributed in S different species. Then Fisher showed on quite plausible assumptions (Fisher *et al.* 1943) that the expected number of species with i individuals would be proportional to x^i/i where x is a parameter, having a value between 0 and 1, which depends on N and S. If the factor of proportionality is α then the expected numbers with 1, 2, 3, etc., individuals per species are αx, $\alpha x^2/2$, $\alpha x^3/3$, ... These are the individual terms in the series expansion of $-\alpha \log(1-x)$ and it is for this reason that the series is known as the logarithmic series. The total number of individuals is $\alpha x + \alpha x^2 + \alpha x^3 + \ldots = \alpha x/(1-x)$. To estimate α and x we thus have two equations

$$\left.\begin{aligned} S &= -\alpha \log(1-x), \\ N &= \frac{\alpha x}{1-x}. \end{aligned}\right\} \qquad (1)$$

It has often been found that, using formulae (1) to estimate α and x for a set of data, the logarithmic series gives a good fit to the observations. Williams (1960) postulates that it might not be unreasonable to use this series to fit to the number of insects and insect species in the world, although a better fit might be obtained by using a truncated lognormal distribution. As rough estimates he takes $N = 10^{18}$ and $S = 3 \times 10^6$ and with these values calculates $x = 1 - 1{\cdot}002 \times 10^{-13}$ and $\alpha = 1{\cdot}002 \times 10^5$. He then asks how many species there are with more than R individuals. This quantity is

$$\alpha S(x, R) = \sum_{r=R+1}^{\infty} \alpha x^r / r.$$

The remainder of this article is concerned with the numerical evaluation of $S(x, R)$ particularly for large values of R.

Evaluation of the remainder $S(x, R)$

An exact formula for the remainder term, true for all values of x and R is

$$S(x, R) = \log(1-x) - \sum_{1}^{R} 1/n - \int_0^{1-x} \frac{(1-\phi)^R - 1}{\phi}\, d\phi. \qquad (2)$$

To evaluate this formula for large values of R we deal first with the term $\sum_1^R 1/n$. This has been tabulated by Glover (1930) for $R = 2(1)450$. If these tables are not available or if larger values of R are of interest recourse can be had to the well-known formula

$$\sum_1^R 1/n = \gamma + \tfrac{1}{2}R + \log R - U_R, \tag{3}$$

where $\gamma = 0{\cdot}5772156649$ is Euler's constant and

$$U_R = \frac{B_1}{2R^2} - \frac{B_2}{4R^4} + \frac{B_3}{6R^6} - \ldots,$$

the B_i being Bernoulli numbers. The series U_R rapidly approaches zero as R increases, specimen values being given in Table 1.

Table 1

R	U_R
10	0·00083 25039
50	·00003 33320
100	·00000 83333
200	·00000 20833
500	·00000 03333
1,000	·00000 00833
2,000	·00000 00208
5,000	·00000 00033
10,000	·00000 00008

It is thus permissible to replace $\sum_1^R 1/n$ by

$$\gamma + \tfrac{1}{2}R + \log R. \tag{4}$$

If $R > 100$ the result is at most one figure out in the fifth decimal place.

Returning to (2) we now try to find an expression for the integral term more amenable to calculation. A well-known inequality is

$$e^{-t} \geqslant (1-t/n)^n \geqslant e^{-t-t^2/(2n)} \quad \text{if} \quad 0 \leqslant t \leqslant n.$$

Writing $n = R$ and $t = \phi R$, subtracting one, dividing each expression by ϕ, integrating between 0 and $(1-x)$ and noting that

$$\int_0^{1-x} \frac{e^{-R\phi} - 1}{\phi}\, d\phi = -\mathrm{Ei}\,[-R(1-x)] + \log R(1-x) + \gamma \tag{5}$$

we find that

$$-\mathrm{Ei}\,[-R(1-x)] + \log[R(1-x)] + \gamma \geqslant \int_0^{1-x} \frac{(1-\phi)^R - 1}{\phi}\, d\phi$$
$$\geqslant -\mathrm{Ei}\,[-R(1-x)] + \log[R(1-x)] + \gamma - \tfrac{1}{4}R(1-x)^2.$$

Here $-\mathrm{Ei}\,(-u)$ is the exponential integral and has been tabulated by several authors, the most extensive tabulation being that of the New York W.P.A. (1940) where $u = 0(0{\cdot}001)10(0{\cdot}1)15$. Results for some higher values of u are given by Akahiro (1929) where $u = 20(0{\cdot}02)50$. We may thus write approximately

$$\int_0^{1-x} \frac{(1-\phi)^R - 1}{\phi}\, d\phi = -\mathrm{Ei}\,[R(1-x)] + \log R(1-x) + \gamma, \tag{6}$$

the error being less than $\frac{1}{4}R(1-x)^2$. Combining (2), (3) and (6)

$$S(x, R) = -\mathrm{Ei}\,[-R(1-x)] - \frac{1}{2R} + U_R + V, \tag{7}$$

where V is an error term which is less than $\frac{1}{4}R(1-x)^2$. A further simplification can be made when $R(1-x)$ is small. In this case

$$\int_0^{1-x} \frac{(1-\phi)^R - 1}{\phi}\, d\phi = -R(1-x) \tag{8}$$

with an error less than $\frac{1}{8}R^2(1-x)^2$. Thus we have

$$S(x, R) = -\log(1-x) - \sum_1^R 1/n + R(1-x) + RV. \tag{9}$$

An alternative approach, due to P. M. Grundy (unpublished work) is to express $\int_R^\infty \frac{X^u}{u}\,du$ in terms of its Euler–Maclaurin expansion (Jeffreys & Jeffreys, 1950, pp. 278–83). This gives

$$S(x, R-1) = \mathrm{Ei}\,[R\log x] + x^R/2R \qquad (10)$$

with an error less than $(1 - R\log x)\,x^R/(12R^2)$. At first sight this seems a neater result than those obtained above, and in fact for many values of x and R it is. Unfortunately, for very large values of R and x near to unity, x^R must for ease of computation be replaced by $e^{-R(1-x)}$ and an investigation of the error involved in this approximation would invoke inequalities such as the one used previously. In fact, making this approximation, (10) can be shown to be equivalent to (7).

The use of the formulae

The formulae appear at first sight to be rather complex, but in fact little difficulty should be found in using them. A summary of the situations when the different formulae should be used is given below.

(1) $R^2(1-x)^2$ *is small.* Use formula (9). The error is less than $\frac{1}{3}R^2(1-x)^3$. If tables of $\sum_1^R 1/n$ are not available, or if R is too large formula (3) may be used to evaluate $\sum_1^R 1/n$.

(2) $R(1-x)^2$ *is small.* If $R(1-x)$ is appreciable but $R(1-x)^2$ is small use formula (7). The error is less than $\frac{1}{3}R(1-x)^2$.

(3) If x is not too small use Grundy's formula (10).

These formulae should be sufficient for most calculations.

Table 2. *Calculation of the remainder term $S(x, R)$ for 3×10^6 $(= S)$ species and $x = 1 - 1{\cdot}002\times10^{-13}$*

R	$S(x,R)$	$100\left\{1+\frac{S(x,R)}{\log(1-x)}\right\}$	Notes	Method
1	28·932	3·3	—	Direct summation
10	27·003	9·8	Formulae (9) and (2) (with the term $1/(2R)$) gives 27·002	Direct summation
10^2	24·744	17·3	Direct summation gives the same results	Formulae (9) and (2), including the term $1/(2R)$
10^3	22·446	25·0	Direct summation gives the same results	Formulae (9) and (2), including the term $1/(2R)$
10^4	20·144	32·7	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^5	17·842	40·4	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^6	15·539	48·1	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^7	13·236	55·8	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^8	10·934	63·5	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^9	8·631	71·2	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^{10}	6·330	78·9	—	Formulae (9) and (2) (the term $1/(2R)$ is negligible)
10^{11}	4·038	86·5	These values agree with those obtained by formulae (2) and (9)	Formula (7)
10^{12}	1·823	93·9	These values agree with those obtained by formulae (2) and (9)	Formula (7)
10^{13}	0·219	99·3	—	Formula (7)
10^{14}	0·000	100·0	—	Formula (7)

As an example Williams (1960) suggested values of $N = 10^{18}$ and $S = 3\times10^6$ for the number of insects and insect species in the world and calculates $x = 1 - 1{\cdot}002\,(10^{-13})$, $\alpha = 1{\cdot}002\,(10^5)$. With these figures and the formulae derived above Table 2 is produced. This agrees substantially with Williams's Table 4 (see the columns for $S = 3\times10^6$) and extends his table for the higher values of R. If the other two columns of Williams's Table 4 are extended (i.e. $S = 2\times10^6$ and $S = 5\times10^6$), similar sets of results are obtained and the main conclusion, that the percentage number of species with more than a specified number of individuals per species remains fairly steady over the range of values of S considered, remains true for the higher values of R.

The median

The median of the logarithmic series is defined as the value of R satisfying the equation

$$S(x, R) = -\tfrac{1}{2}\log(1-x). \tag{11}$$

If R is large (> 100) and $R(1-x)$ is small, then (9) becomes $S(x, R) = -\log R(1-x) - \log R - \gamma$ so that the median is given by $-\frac{1}{2}\log(1-x) = \log R + \gamma$, i.e.

$$R = \frac{e^{-\gamma}}{\sqrt{(1-x)}} = \frac{0{\cdot}56146}{\sqrt{(1-x)}}. \tag{12}$$

This, apart from a term 0·81524, is the formula given on page 144 of Williams (1960) and attributed to P. M. Grundy. The 0·81524 given should there occur as an additive term and not a multiplicative one as printed.

A further stage of approximation adds in the Grundy terms $(\frac{1}{2} + e^{-2\gamma})$, but for high values of S this is of course quite trivial. In many ways the best approach appears to be to produce a table similar to Table 2 above and then interpolate to find the value of $R-1$ corresponding to 50 %. For Table 2 this would occur between $R = 10^6$ and $R = 10^7$ and where the curve of $S(x, R)$ is linear relative to a log-scale of R. Noting that $-\log(1-x) = 29{\cdot}932$, linear interpolation gives

$$15{\cdot}539 - 13{\cdot}236 = \frac{15{\cdot}539 - \frac{1}{2}(29{\cdot}932)}{7 - \log R},$$

$\log R = 6{\cdot}2488$, $R = 1773000$, the value obtained by formula (12). If the median occurs on the curved portion of the $S(x, R)$ curve then some more elaborate form of inverse interpolation using, say, four adjacent points on the $S(x, R)$ curve, will have to be used.

References

Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

Williams, C. B. (1960). The range and pattern of insect abundance. *Amer. Nat.* **94**, 137–51.

Glover, J. W. (1930). *Tables of Applied Mathematics in Finance, Insurance and Statistics.* Ann Arbor, Michigan: Wahr.

New York W.P.A. (1940). *Tables of Sine, Cosine and Exponential Integrals*, **1** and **2**.

Akahiro, T. (1929). *Sci. Pap. Inst. Phys. Chem. Res. Tokyo.* Table no. 3, pp. 181–215.

Jeffreys, H. & Jeffreys, B. S. (1950). *Mathematical Physics* (second edition). Cambridge University Press.

On a property of balanced designs

By M. Atiqullah

Birkbeck College, London

1. Introduction and summary

Special classes of balanced designs are well known and widely used. In this paper, a necessary and sufficient condition for a general class of connected designs to be balanced is derived. This is a natural extension of a result of Tocher (1952) and of Thompson (1956), and it appears to be simpler than the generalization given by Rao (1958). A simple expression for calculating the efficiency factor for a connected balanced design is obtained.

Fisher (1940) established that $b \geqslant v$ for a balanced incomplete block design with v treatments and b blocks. This inequality is shown to be true for a wider class of designs, similar to the balanced incomplete block designs but with blocks of different sizes.

2. Notation and preliminary results

Consider v treatments arranged in b blocks in a design whose incidence matrix is $\mathbf{N} = (n_{ij})$, where n_{ij} denotes the number of experimental units in the ith block getting the jth treatment. The ith block is of size k_i $(i = 1, 2, \ldots, b)$ and the jth treatmen tis replicated r_j times $(j = 1, 2, \ldots, v)$. These may be

conveniently written as the diagonal matrices $\mathbf{K} = \text{diag}(k_1, \ldots, k_b)$, $\mathbf{R} = \text{diag}(r_1, \ldots, r_v)$. The total yield of the jth treatment is T_j and that of the ith block is B_i. On writing $\mathbf{T}' = (T_1, \ldots, T_v)$ and $\mathbf{B}' = (B_1, \ldots, B_b)$ in matrix notation, the prime denoting transposition, the adjusted normal equations for estimating the vector of treatment constants $\mathbf{t}$ can be written under the usual asumptions as

$$\mathbf{C}\hat{\mathbf{t}} = \mathbf{Q}, \tag{1}$$

where $$\mathbf{Q} = \mathbf{T} - \mathbf{N}'\mathbf{K}^{-1}\mathbf{B}, \quad \mathbf{C} = \mathbf{R} - \mathbf{N}'\mathbf{K}^{-1}\mathbf{N}.$$

The dispersion matrix of $\mathbf{Q}$ is known to be $\sigma^2\mathbf{C}$, where σ^2 is the intra-block error. Since each row (or column) of $\mathbf{C}$ adds up to zero, the rank of $\mathbf{C}$ is at most $v-1$, and $\mathbf{e}' = (v^{-\frac{1}{2}}, v^{-\frac{1}{2}}, \ldots, v^{-\frac{1}{2}})$ is the latent vector corresponding to the zero root. If the rank of $\mathbf{C}$ is $v-1$, the design is said to be connected (Bose, 1950), and in that case the solution of (1) is given by

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{\Gamma}_1 \mathbf{U}^{-1} \boldsymbol{\Gamma}_1')\,\mathbf{Q}. \tag{2}$$

Here $\boldsymbol{\Gamma} = (\mathbf{e} : \boldsymbol{\Gamma}_1)$ is an orthogonal matrix transforming $\mathbf{C}$ into a diagonal matrix $\mathbf{u} = \text{diag}(u_1, u_2, \ldots, u_{v-1})$, whose elements are the non-zero latent roots of $\mathbf{C}$, and $\hat{\tau}_j = \hat{t}_j - \bar{t}$, $(\bar{t} = (1/v)\sum_j \hat{t}_j)$. The dispersion matrix of $\boldsymbol{\tau}$ follows from (2) as $\sigma^2\mathbf{D}$, where

$$\mathbf{D} = \boldsymbol{\Gamma}_1 \mathbf{u}^{-1} \boldsymbol{\Gamma}_1', \tag{3}$$

and each row (or column) of $\mathbf{D}$ adds up to zero.

3. Condition for a design to be balanced

A design is said to be balanced if every elementary contrast $\tau_j - \tau_{j'}$ is estimated with the same variance. We have the following theorem.

Theorem 1. A necessary and sufficient condition for a connected design to be balanced is that every treatment constant is estimated with the same variance and every pair of treatment constants with the same covariance.

To prove the necessary condition, we can write, following Kempthorne (1956), the average variance for elementary contrasts, $\hat{\tau}_j - \hat{\tau}_{j'}$, as

$$\frac{1}{v(v-1)} \sum_{j \neq j'} \text{var}\,(\hat{\tau}_j - \hat{\tau}_{j'}) = \frac{2\sigma^2}{H}, \tag{4}$$

where H is the harmonic mean of the non-zero latent roots of $\mathbf{C}$.

Since the trace of the matrix $\mathbf{D}$ is equal to the sum of its latent roots, we get

$$\sum_j \text{var}\,(\hat{\tau}_j) = \frac{\sigma^2}{H}(v-1). \tag{5}$$

Also, $$\text{var}\,(\hat{\tau}_1 - \hat{\tau}_2) + \text{var}\,(\hat{\tau}_1 - \hat{\tau}_3) + \ldots + \text{var}\,(\hat{\tau}_1 - \hat{\tau}_v) = v\,\text{var}\,(\hat{\tau}_1) + \sum_j \text{var}\,(\hat{\tau}_j), \tag{6}$$

$$\text{var}\,(\hat{\tau}_2 - \hat{\tau}_1) + \text{var}\,(\hat{\tau}_2 - \hat{\tau}_3) + \ldots + \text{var}\,(\hat{\tau}_2 - \hat{\tau}_v) = v\,\text{var}\,(\hat{\tau}_2) + \sum_j \text{var}\,(\tau_j), \tag{7}$$

for the rows of $\mathbf{D}$ add up to zero.

Clearly, $\text{var}\,(\hat{\tau}_1) = \text{var}\,(\hat{\tau}_2)$ if the design is balanced (i.e. if every elementary contrast, $\tau_j - \tau_{j'}$, is estimated with the same variance).

Hence, it follows from (5) that

$$\text{var}\,(\hat{\tau}_j) = \frac{\sigma^2}{H}\left(1 - \frac{1}{v}\right) \tag{8}$$

and the relation (4) gives

$$\text{var}\,(\hat{\tau}_j - \hat{\tau}_{j'}) = \frac{2\sigma^2}{H}, \tag{9}$$

or $$2\,\text{var}\,(\hat{\tau}_j) - 2\,\text{cov}\,(\hat{\tau}_j, \hat{\tau}_{j'}) = \frac{2\sigma^2}{H}, \tag{10}$$

or $$\text{cov}\,(\hat{\tau}_j, \hat{\tau}_{j'}) = -\frac{\sigma^2}{Hv}. \tag{11}$$

For the proof of sufficiency, note that if $\text{var}\,(\hat{\tau}_j)$ and $\text{cov}\,(\hat{\tau}_j, \hat{\tau}_{j'})$ are independent of j and j', so obviously is $\text{var}\,(\hat{\tau}_j) + \text{var}\,(\hat{\tau}_{j'}) - 2\,\text{cov}\,(\hat{\tau}_j, \hat{\tau}_{j'})$.

4. Efficiency factor for a connected balanced design

If a connected design is balanced, the non-zero latent roots of $\mathbf{C}$ are identical, $u_1 = u_2 = \ldots = u_{v-1} = H$, and we have from (2) and (3) that

$$\hat{\boldsymbol{\tau}} = \frac{1}{H}(\mathbf{I}_v - \mathbf{e}\mathbf{e}')\,\mathbf{Q} = \frac{1}{H}\mathbf{Q}, \tag{12}$$

$$\mathbf{D} = \frac{1}{H}(\mathbf{I}_v - \mathbf{e}\mathbf{e}'). \tag{13}$$

Hence
$$\operatorname{cov}(\mathbf{Q}, \mathbf{Q}') = H^2 \operatorname{cov}(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\tau}}'), \tag{14}$$

or
$$\mathbf{C} = \mathbf{H}\mathbf{D}, \tag{15}$$

or
$$\mathbf{R} - \mathbf{N}'\mathbf{K}^{-1}\mathbf{N} = H\mathbf{D} = H\begin{pmatrix} 1-\frac{1}{v} & -\frac{1}{v} & \ldots & -\frac{1}{v} \\ -\frac{1}{v} & 1-\frac{1}{v} & \ldots & -\frac{1}{v} \\ -\frac{1}{v} & -\frac{1}{v} & \ldots & 1-\frac{1}{v} \end{pmatrix}. \tag{16}$$

Since $\operatorname{trace}(\mathbf{N}'\mathbf{K}^{-1}\mathbf{N}) = \operatorname{trace}(\mathbf{R} - \mathbf{H}\mathbf{D})$, we have

$$\sum_i \frac{1}{k_i}\sum_j n_{ij}^2 = \sum_j r_j - H(v-1), \tag{17}$$

or
$$H = \frac{\sum_j r_j - \sum_i \frac{1}{k_i}\sum_j n_{ij}^2}{v-1}. \tag{18}$$

Hence, the efficiency factor (E.F.) relative to a randomized block with the same number of experimental units n $(= \sum_j r_j = \sum_i k_i)$ is given by

$$\text{E.F.} = \frac{H}{n/v} = \frac{\sum_j r_j - \sum_i \frac{1}{k_i}\sum_j n_{ij}^2}{n\{1-(1/v)\}}. \tag{19}$$

We call a design binary if no treatment occurs more than once in the same block, i.e. if $n_{ij} = 0$ or 1. Then $n_{ij}^2 = n_{ij}$ and for a connected binary balanced design we have

$$H = \frac{n-b}{v-1}, \tag{20}$$

$$\text{E.F.} = \frac{n-b}{n\{1-(1/v)\}}. \tag{21}$$

Further, if the design be equi-replicate,

$$\text{E.F.} = \frac{vr-b}{r(v-1)}, \tag{22}$$

where $r_1 = r_2 = \ldots = r_v = r$.

5. Extension of Fisher's inequality

Theorem 2. For a connected binary equi-replicate balanced design, b cannot be less than v.

On substituting the value of $H = (vr-b)/(v-1)$ in (16) we get

$$\mathbf{N}'\mathbf{K}^{-1}\mathbf{N} = \begin{pmatrix} \frac{b}{v} & \frac{vr-b}{v(v-1)} & \cdots & \frac{vr-b}{v(v-1)} \\ \frac{vr-b}{v(v-1)} & \frac{b}{v} & \cdots & \frac{vr-b}{v(v-1)} \\ \vdots & \vdots & & \vdots \\ \frac{vr-b}{v(v-1)} & \frac{vr-b}{v(v-1)} & \cdots & \frac{b}{v} \end{pmatrix}. \tag{23}$$

Hence, the determinant

$$|\mathbf{N}'\mathbf{K}^{-1}\mathbf{N}| = \left(\frac{b-r}{v-1}\right)^{v-1} r. \tag{24}$$

Suppose $v > b$. Then $|\mathbf{N}'\mathbf{K}^{-1}\mathbf{N}| = 0$ and consequently $b = r$, which contradicts the fact that the design is binary. Hence $v \leqslant b$.

It is interesting to note that for $v = b$ and consequently $r = (k_1 + k_2 + \ldots + k_v)/v$, we have the determinantal value from (24) as

$$|\mathbf{N}'\mathbf{N}| = (k_1 k_2 \ldots k_v)\left(\frac{v-r}{v-1}\right)^{v-1} r \tag{25}$$

$$\leqslant r^{v+1}\left(\frac{v-r}{v-1}\right)^{v-1}. \tag{26}$$

Since the left-hand side of (25) is an integral square, so also must be the right-hand side. Hence (25) can be used as a test for the possible existence of a connected binary equi-replicate design having specified values of $v = b, k_1, \ldots, k_v$.

I am indebted to Dr D. R. Cox and the referee for suggesting a number of improvements to the original draft.

References

Bose, R. C. (1950). *Least Squares Aspects of Analysis of Variance.* Institute of Statistics, University of North Carolina.

Fisher, R. A. (1940). An examination of the different possible solutions of a problem in incomplete blocks. *Ann. Eugen., Lond.*, **10**, 52–75.

Kempthorne, O. (1956). The efficiency factor of an incomplete block design. *Ann. Math. Statist.* **27**, 846–9.

Rao, V. R. (1958). A note on balanced designs. *Ann. Math. Statist.* **29**, 290–4.

Roy, J. (1958). On the efficiency factor of block designs. *Sankhyā*, **19**, 181–8.

Thompson, W. A. (1956). A note on balanced incomplete block designs. *Ann. Math. Statist.* **27**, 842–6.

Tocher, K. D. (1952). The design and analysis of block experiments. *J. R. Statist. Soc.* B, **14**, 45–91.

Aliasing in partially confounded factorial experiments

By M. J. R. Healy and J. C. Gower

Rothamsted Experimental Station

In a factorial experiment with factors at p levels with $p > 2$, the confounding structure is usually determined by breaking down the interactions into groups of $(p-1)$ contrasts by means of systems of linear congruences (see, for example, Kempthorne, 1952). For example, to confound a 3^3 experiment in blocks of 9 units each, we may allot each treatment to block 1, 2 or 3 according as

$$(\text{level of 1st factor}) + (\text{level of 2nd factor}) + 2 \times (\text{level of 3rd factor}) = 0, 1 \text{ or } 2 \pmod 3.$$

This procedure will confound a pair of degrees of freedom belonging to the 3-factor interaction which may be denoted by ABC^2. The other 6 d.f. from this interaction are similarly denoted in pairs by AB^2C, AB^2C^2 and ABC, respectively. This type of subdivision has no particular meaning in the interpretation of the results.

Particularly when the factor levels are quantitative, a different kind of breakdown is often of interest. For example, each main effect can be broken down into $(p-1)$ orthogonal polynomial contrasts, and these induce corresponding breakdowns of the interactions. When an experiment is confounded using one set of contrasts and analysed using another, certain difficulties arise which are closely analogous to the problems of aliasing in fractionally replicated experiments. This fact was originally pointed out by Yates (1937, § 12*c*) but does not seem to have attracted much attention since that time.

The least squares estimates of a set of treatment contrasts can readily be obtained. Consider an experiment with r replicates of t treatments carried out in b blocks of k units each. Its design can be specified by the incidence matrix $\mathbf{N}$, of dimensions $t \times b$. The rows of $\mathbf{N}$ correspond to the treatments and the columns to the blocks; $N_{ij} = 1$ if treatment i occurs in block j, otherwise $N_{ij} = 0$. The normal equations can be written

$$\begin{bmatrix} r\mathbf{I} & \mathbf{N} \\ \mathbf{N}^T & k\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{T} \\ \mathbf{B} \end{bmatrix},$$

where $\mathbf{t}$ and $\mathbf{b}$ are the vectors of treatment and block constants, and $\mathbf{T}$ and $\mathbf{B}$ the vectors of treatment and block totals. Eliminating $\mathbf{b}$, we find that the normal equations for $\mathbf{t}$ can be written

$$\left(r\mathbf{I}-\frac{1}{k}\mathbf{N}\mathbf{N}^T\right)\mathbf{t}=\mathbf{Q},$$

where $\mathbf{Q}=\mathbf{T}-(1/k)\,\mathbf{N}\mathbf{B}$ is the vector of treatment totals of deviations from block means.

Suppose next that an orthogonal set of treatment contrasts $\mathbf{v}$ is defined by

$$\mathbf{v}=\mathbf{G}\mathbf{t},$$

where $\mathbf{G}$ is an orthonormal matrix, the elements in the first row being all equal. Then the normal equations for $\mathbf{v}$ are easily shown to be

$$\left(r\mathbf{I}-\frac{1}{k}\mathbf{M}\mathbf{M}^T\right)\mathbf{v}=\mathbf{G}\mathbf{Q},$$

where $\mathbf{M}=\mathbf{G}\mathbf{N}$. Difficulties in analysis and interpretation may arise when the matrix on the left-hand side of this equation is not diagonal.

As a simple example consider a single replicate 3^2 experiment in blocks of 3 units, confounding AB^2. The incidence matrix is

Treatment			
00	1	0	0
01	0	1	0
02	0	0	1
10	0	0	1
11	1	0	0
12	0	1	0
20	0	1	0
21	0	0	1
22	1	0	0

To analyse this in terms of linear and quadratic components, we take $\mathbf{G}$ to be $\mathbf{C}\mathbf{G}_0$ where $\mathbf{G}_0$ is equal to

A_L	1	1	1	1	1	1	1	1	1
A_Q	−1	−1	−1	0	0	0	1	1	1
B_L	−1	−1	−1	2	2	2	−1	−1	−1
B_Q	−1	0	1	−1	0	1	−1	0	1
A_LB_L	−1	2	−1	−1	2	−1	−1	2	−1
A_LB_Q	1	0	−1	0	0	0	−1	0	1
A_QB_L	1	−2	1	0	0	0	−1	2	−1
A_QB_Q	1	0	−1	−2	0	2	1	0	−1
	1	−2	1	−2	4	−2	1	−2	1

and $\mathbf{C}$ is a diagonal matrix that normalizes the rows.

The information matrix for these contrasts is

$$r\mathbf{I}-\frac{1}{k}\mathbf{M}\mathbf{M}^T=\begin{bmatrix}0&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&\frac{1}{2}&0&0&-\frac{1}{2}\\0&0&0&0&0&0&\frac{1}{2}&\frac{1}{2}&0\\0&0&0&0&0&0&\frac{1}{2}&\frac{1}{2}&0\\0&0&0&0&0&-\frac{1}{2}&0&0&\frac{1}{2}\end{bmatrix}.$$

From this it will be seen that A_LB_L is completely aliased with A_QB_Q, and A_LB_Q with A_QB_L. These contrasts can be estimated only if some linear relationship between the members of the pair, with known coefficients, is assumed to hold.

If two replicates of the experiment are used, with AB confounded in one and AB^2 in the other, balance is attained and this type of difficulty does not arise. For, in terms of the contrasts used in the confounding, each section of the information matrix that corresponds to interactions of a given order will be simply a multiple of the unit matrix, and this property is unchanged when a different breakdown of the interactions is imposed.

Certain common designs will now be examined in terms of orthogonal polynomial contrasts.

1. 3 × 3 × 3 in blocks of 9

There are four types of replicate, each confounding a pair of d.f. from the 3-factor interaction. If a single replicate is used (confounding, say, ABC), the eight polynomial-type contrasts fall into two sets of four, such that the contrasts within each set are aliased together. With two or three replicates, confounding a different pair of d.f. in each, the information matrix is no longer singular, but it still contains off-diagonal terms so that explicit matrix inversion is necessary for complete solution of the normal equations. The resulting estimates are correlated, and the situation may be referred to as *partial aliasing*. With four replicates, balance is obtained and all the estimates can be found by simply dividing the appropriate Q contrast totals by 3, the relative information ($\frac{3}{4}$) times the number of replicates.

2. 3 × 2 × 2 in blocks of 6

Yates (1937) gives a balanced design in 3 replicates. Because the confounding is balanced, the polynomial contrasts can be obtained from the contrast totals of the Q's by dividing by three times the relative information.

3. 3 × 3 × 2 in blocks of 6

Here balance requires four replicates. If only two replicates are used, partial aliasing occurs in both the AB and ABC interactions. In particular, $A_L B_L$ is partially aliased with $A_Q B_Q$ and $A_L B_Q$ with $A_Q B_L$.

The occurrence of partial and total aliasing is of importance when considering the analysis of factorial experiments on electronic computers. Thus Tocher (discussion in Yates, Healy & Lipton, 1957), has suggested an extension of Yates's 'plus-and-minus' technique to factors at more than two levels—in essence, the main effects are broken down into single degrees of freedom and the matrix **G** is then built up as a direct product of smaller matrices, providing a rapid and compact scheme for the formation of the vector **GQ**. However, it is still necessary to invert the information matrix, and when this may be no longer diagonal a fully generally program is likely to become cumbersome.

References

Kempthorne, O. (1952). *The Design and Analysis of Experiments.* New York: John Wiley and Sons Inc.

Yates, F. (1937). The design and analysis of factorial experiments. *Tech. Commun. Emp. Bur. Soil Sci.* **35**.

Yates, F., Healy, M. J. R. & Lipton, S. (1957). Routine analysis of replicated experiments on an electronic computer. *J. R. Statist. Soc.* B, **19**, 234–54.

Studies in the History of Probability and Statistics. XII. The Book of Fate

By M. G. Kendall

Research Techniques Division, London School of Economics

1. Divination by chance mechanism is a very ancient practice and is reported from many countries: China, Tibet, Greece, Rome and among the Germanic tribes known to the ancient world. Sticks and dice were both in use. This, for example, is Tacitus's account of the Germani:

'To divination and the drawing of lots (*auspices sortesque*) they pay as much attention as anyone. The method of drawing lots is standardized. They take a bough from some nut tree and cut it into strips, on which certain runes are written. These strips are scattered at random on a white cloth. Then the priest (if it is a public occasion) or the head of the household (if it is a family matter) prays to the gods, turns his eyes to heaven and take up three sticks, one at a time. The runes so collected are then interpreted.'

2. This is one of the earliest instances I know of divination in Europe by reference to a previously prepared script. The Romans also had their Sybilline Books. It was perhaps more common in Rome, and not uncommon among the Germanic tribes, to divine from natural phenomena such as the voices of birds or the entrails of a sacrifice. Mantic practices of all kinds flourished under the later Roman Empire until Constantine's time, when the official adoption of Christianity put a stop to the investigation of God's will by lot. Occasional attempts at revival were discouraged. Three medieval prelates who

were each claiming possession of the bones of a saint decided, in what nowadays would probably be regarded as a very sporting spirit, to settle the dispute by drawing lots. But they were reprimanded for doing so and their agreement was cancelled by higher authority.

3. Vestiges of the religious practices of one age provide the superstitions of the next and the pastimes of the one after that. The Gods were transformed into demons and then into fairies; their rites become customs; divination dwindled into fortune telling. But, although consulting the oracle may have become only an entertainment it was taken seriously enough to survive every effort to suppress it. There does not exist much documentary evidence of the practice before the advent of printing but occasional references are enough to show that it never disappeared. In an earlier article (1956) I have referred to some of the medieval poems like the 'Chaunce of the Dyse' which told fortunes or character by dice throwing. The so-called *sortes Virgilianae* seem also to have been practised over the period among those who were lucky enough to have books. A volume of Virgil was opened at random, a passage also chosen at random and then applied to the situation requiring elucidation. (Virgil in the dark ages was regarded as a notable necromancer.) Some comments of striking aptitude are reported but the only example I can quote is modern, probably apocryphal and illustrates the depths of degradation to which Sortilege has descended: when Margaret Bondfield, as Minister of Labour, was about to move a resolution requiring the renewal of unemployment assistance, known at the time as 'the dole' the *sortes* are said to have yielded the famous line from the Aeneid:

'Infandum, regina, jubes renovare dolorem.'

4. The advent of playing cards and of printing revived interest in this type of entertainment. In 1540 Francesco Marcolini of Forli published *Le Sorti*, a large and elaborately illustrated volume for consulting the oracle by using a pack of playing cards. This must have been a fairly popular work, for a second edition appeared in 1550. The scope and elaboration in this book suggests that there may somewhere be earlier prototypes. Marcolini has 50 questions, 13 for men, 13 for women and 24 for men and women. To each question he has 90 answers written in rhymed triplets, 4500 answers in all. The sortilege consists of going from a question to one of the 90 corresponding answers by an elaborate piece of mystification involving the drawing of cards at three different stages. The questions themselves are an interesting collection of the topics which exercised the minds of the gentlefolk of sixteenth-century Italy, and most of the politer ones would be found in any fortune-telling book of today. For example, question 3 for men is: whether it is better to choose a beautiful or a plain wife (*bella o brutta*). One possible answer to this (to take an example offered in illustration by Marcolini himself) is

'Una brutta e una bella ne torrai
Se vuol la legge; et se no vuol, sei matto
Se una sol moglie e brutta pigliorai.'

This, I take it, means: 'Have one of each if the law permits, but if not, you would be crazy to choose a plain one.'

5. For students of combinatorial analysis and experimental design there is some mild interest in the mechanism by which Marcolini gets from question to answer. He works with a pack from which four cards of each suit, 3, 4, 5, 6 have been stripped. In point of fact, he makes no distinction between suits, so he might just as well work with a pack of nine cards of one suit. Two are chosen (effectively with replacement) so that there are 45 possible pairs (the order of a pair being immaterial). These 45 are displayed in full on a single page, one page to each of the 50 questions. The 45 possibilities, however, are divided into five groups of nine so that effectively there are only five possibilities per question. Marcolini then adds two more stages each involving choice of one card. But he again reduces the possibilities so that effectively there are only $5 \times 18 = 90$ per question. How or why these numbers were chosen is a mystery. Personally I do not feel much confidence in Marcolini's command of combinatorial analysis. There appear to me to be places where he failed to preserve the balance of his incomplete blocks and just faked them. But it is easy to be superior with the advantage of four centuries' experience, and his drawings are fascinating.

6. In *Le Sorte* we have the last stage of degeneration of the old sortilege, a deliberate obfuscation of the selective process. Henceforward the Book of Fate is of no scientific interest. It still exists. I recently bought a publication called *Napoleon's Book of Fate* part of which is evidently a lineal descendant of *Le Sorte*.

7. One curious development in the musical field may be worth recording. In 1757 a certain Johann Phillip Kirnberger published in Berlin a book which enabled Polonaises and Minuets to be composed by throwing dice. What is more, they were composed for two violins and pianoforte. This game seems to have caught on. Haydn is credited with a *Guioco Filarmonico* published in 1790 for the composition of

minuets. To C.P.E. Bach has been attributed a publication of the same year for composing waltzes. In 1793, after Mozart's death, there was published (in four languages) a similar method, of composing with two dice. This work is attributed to Mozart in the Köchel–Einstein index (Anhang 294[d]). It reappeared in 1806 in England (C. Wheatstone, London) as '*Mozart's Musical Game*, fitted in an elegant box, showing by an easy system to compose an unlimited number of Waltzes, Rondos, Hornpipes and Reels.' I have not been able to trace the Wheatstone edition or to find out what was in the elegant box, but the German edition and Kirnberger's work are both in the British Musum.

These works are prefaced by fairly lengthy descriptions of how to apply the tables which are given. They give no account of the method of construction. I am certain that the general method in all these works must have been the same. We write down a simple harmonic sequence of, say, eight chords, finishing on the tonic. We can then compose a large number of bars from each of these chords by using notes of the triads or their inversions. We choose eleven of them and number them from 2 to 12 corresponding to the total points on two dice; and hence we can pick out one at random.

Likewise we can do the same for each of the eight chords in the sequence. And if we put the resulting eight bars together *in the right order* we shall, with a little luck, have an acceptable air. The same procedure can be followed with the left hand.

There only remains to put in the mystique. This is done by arranging our 88 bars in random order and then providing a table to unscramble them.

8. The student of probability and statistics will, I fear, derive little of relevance to his subject in these fortune-telling books and games. The chance mechanisms are of interest, perhaps, but there is no sign that, in any of the cases I have described, there was any appreciation of relative frequency.

References

Marcolini, F. (1540). *Le Sorte*. Venetia.

Kendall, M. G. (1956). Studies in the history of probability and statistics. II. The beginnings of a probability calculus. *Biometrika*, **43**, 1.

Kirnberger, J. P. (1757). *Der allerzeit fertige Polonaisen und Menuettencomponist*. Berlin: G. L. Winter.

Köchel, L. von (1937). *Chronologisch-theoretisches Verzeichnis Sämtlicher Tonwerke W. A. Mozarts*. Dritte Auflage, bearbeitet von Alfred Einstein. Leipzig: Breitkopf und Härtel.

A derivation of the Borel distribution

By J. C. TANNER

Road Research Laboratory, Department of Scientific and Industrial Research

In a queueing process with random arrivals at a rate q per unit time and constant service time β, the number of units served during a busy period follows a Borel distribution

$$p_{n,1} = \frac{e^{-n\beta q}(n\beta q)^{n-1}}{n!} \quad (n = 1, 2, \ldots). \tag{1}$$

This was shown by Borel (1942), whose argument was extended by Tanner (1953) to show that the distribution of the number of units served in a busy period starting with an accumulation of r units is

$$p_{n,r} = \frac{e^{-n\beta q}(\beta q)^{n-r} r . n^{n-r-1}}{n-r!} \quad (n = r, r+1, \ldots). \tag{2}$$

The object of this note is to give a simple derivation of the distribution (2).

The fact that the expression (2) may be put into the form

$$p_{n,r} = \frac{e^{-n\beta q}(n\beta q)^{n-r}}{n-r!} \times \frac{r}{n} \quad (n = r, r+1, \ldots) \tag{3}$$

suggests that there may be a simple derivation based on the Poisson distribution. The following are necessary and sufficient conditions for a busy period to contain just n units:

(i) precisely $n-r$ units shall arrive within a time $n\beta$;

(ii) the pattern of these arrivals shall be 'admissible'. The term 'admissible' is used here to mean that at least one unit must arrive during a time $r\beta$, at least two within $(r+1)\beta$, and so on, so that the server will remain continuously occupied for the whole of the time $n\beta$.

The probability of (i) is clearly $\dfrac{e^{-n\beta q}(n\beta q)^{n-r}}{n-r!}$

and it thus remains to show that the probability of the $n-r$ arrivals being admissible is r/n.

Consider any particular arrival pattern; suppose that there are a_1 arrivals in the first service period, a_2 in the second and so on (the timing of the arrivals within each service time is irrelevant). Consider now the set of n arrival patterns obtained by permuting this pattern cyclically. It will be sufficient if we can demonstrate that precisely r of these n patterns are admissible.

Suppose that instead of there being r units waiting initially there are an arbitrary number N, greater than $n+r$, and that the arrival pattern $a_1, a_2, \ldots, a_n$ is repeated in the $(n+1)$th, ..., $2n$th service times. Thus the n cyclical permutations of the arrival pattern are obtained by taking the n service times starting with the 1st, 2nd, ..., nth. Consider how the queue size will vary during these $2n$ service times. At the beginning of the first service time, it will be N (including the unit starting to be served), at the

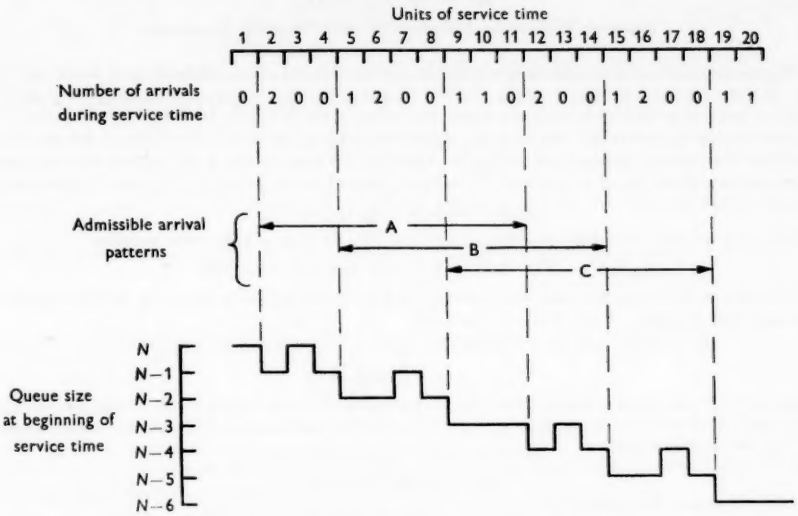
Units of service time
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of arrivals during service time
0 2 0 0 1 2 0 0 1 1 0 2 0 0 1 2 0 0 1 1
Admissible arrival patterns
A
B
C
Queue size at beginning of service time
N
N−1
N−2
N−3
N−4
N−5
N−6

Fig. 1. Queue sizes at the beginning of successive service times, for illustrative example with $n = 10$, $r = 3$.

beginning of the second, $N+a_1-1$ and so on. At the beginning of the $(n+1)$th it will be $N-r$, since there will by that time have been $n-r$ arrivals and n completed services; at the beginning of the $(2n+1)$th it will be $N-2r$. More generally, the queue sizes at the beginnings of the $(n+1)$th, ..., $2n$th service times will be just r less than they were n services times earlier. This means that there must be precisely r service times out of the $(n+1)$th, ..., $2n$th at the start of which the queue size is less than it has previously been. For if the smallest queue size at the beginning of any of the first n service times was $N(k)$ for the kth service time, then the smallest for the next n must be $N(k)-r$ for the $(k+n)$th, and since the queue size can diminish by at most one per service time, there must be $r-1$ out of the service times between the $(n+1)$th and the $(k+n-1)$th inclusive at which the queue sizes decreased for the first time to $N(k)-1, N(k)-2, \ldots, N(k)-r+1$. Also these r service times will be the only ones at the start of which the queue size is less than at the start of any of the preceding n service times, since the queue size at each service time before these n is greater than at at least one of the n.

We can now interpret this result in terms of our original problem. For an arrival pattern is admissible if and only if, starting with r in the queue, the queue size would first drop to zero at the start of the $(n+1)$th service time, in other words if it would be smaller at the start of the $(n+1)$th service time than at the start of any of the first n. Thus there are just r admissible patterns, viz. those which are immediately followed by the queue size in our sequence of $2n$ service times falling for the first time to $N(k)-1, N(k)-2, \ldots, N(k)-r$.

As an illustration, Fig. 1 shows an example with $n = 10$, $r = 3$ and a particular arrival pattern $a_1, a_2, \ldots, a_n$. The admissible arrival patterns are marked A, B and C. In this example, $k = 9, N(k) = N - 3$.

This article is published by permission of the Director of Road Research.

REFERENCES

BOREL, E. (1942). Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au probleme de l'attente à un guichet. *C.R. Acad. Sci., Paris*, **214**, 452–6.
TANNER, J. C. (1953). A problem of interference between two queues. *Biometrika*, **40**, 58–69.

A theorem in trend analysis

BY M. G. KENDALL

Research Techniques Division, London School of Economics

1. The commonest methods of fitting a trend to a series defined at equidistant time-intervals involve the use of some form of moving average. In the standard method a polynomial of order p is fitted to sets of $n\ (= 2m+1)$ points by least squares and the value of the polynomial at the middle point is taken as the trend value at that point. An account of the method is given in volume 2 of my *Advanced Theory of Statistics*. The same trend point is arrived at whether p is even or the next highest odd integer.

In particular the fitting of a quadratic or cubic to sets of seven points gives a moving average with weights $(p = 2, 3, m = 3)$

$$\tfrac{1}{21}[-2, 3, 6, 7, 6, 3, -2]. \tag{1}$$

The fitting of a quartic or quintic to sets of 17 points $(p = 4, 5, m = 8)$ has the weights

$$\tfrac{1}{4199}[195, -195, -260, -117, 135, 415, 660, 825, 883, \ldots]. \tag{2}$$

2. It is more convenient for most purposes to express these formulae in terms of differences. From (1), for example, we have, in an obvious notation,

$$\begin{aligned}\tfrac{1}{21}[-2, 3, 6, 7, \ldots]u_t &= u_t - \tfrac{1}{21}[-2, 3, 6, -14, 6, 3, -2]u_t \\ &= u_t - \tfrac{1}{21}\Delta^4[2, 5, 2].\end{aligned} \tag{3}$$

The fact that the moving average is exact for polynomials of cubic degree implies that the term on the extreme right in (3) must be expressible in terms of fourth differences. We determine the weights by factorizing the polynomial

$$-2x^6 + 3x^5 + 6x^4 - 14x^3 + 6x^2 + 3x - 2$$

into

$$-(x-1)^4(2x^2 + 5x + 2).$$

Likewise (2) can be expressed as

$$u_t + \tfrac{1}{4199}\Delta^6[195, 975, 2665, 5148, 7623, 8778, \ldots]. \tag{4}$$

3. There are several advantages in using these forms. As a rule the different series yield smaller numbers than the original series and the arithmetic is simpler. Likewise, all the terms within the square bracket in expressions like (3) and (4) have the same sign, a result which follows from the later part of this paper as limiting property, but appears to be generally true. Most important of all, the expressions involving differences give us directly the residuals when trend is abstracted from the series.

The moving average [2, 5, 2] in (3) can be represented as [2, 1] [1, 2], and generally any expression of this kind, from its symmetry, can be factorized in the same way. The coefficients in this factorization, however, are not always rational.

4. It will be observed that the weights within square brackets in (3) and (4) are symmetrical unimodal patterns and I proceed to prove a rather remarkable theorem concerning their limiting values as the number of points n and the order p of the fitting increases: in fact, they tend to the ordinates of a normal frequency function taken at equal intervals on either side of the mean.

5. The orthogonal polynomials for equidistant intervals customarily employed in statistics are standardized so that the highest power in x is unity. Thus, if we write

$$P_1 = x - \tfrac{1}{2}(n^2 - 1),$$

the next three polynomials are

$$\begin{aligned}P_2 &= P_1^2 - \tfrac{1}{12}(n^2 - 1), \\ P_3 &= P_1^3 - \tfrac{1}{20}(3n^2 - 7)P_1, \\ P_4 &= P_1^4 - \tfrac{1}{14}(3n^2 - 13)P_1^2 + \tfrac{3}{560}(n^2 - 1)(n^2 - 9).\end{aligned}$$

If we let n be large such that $P_j/n^j = Q_j$ remains finite we have the limiting values

$$Q_2 = Q_1^2 - \tfrac{1}{12},$$
$$Q_3 = Q_1^3 - \tfrac{3}{20}Q_1,$$
$$Q_4 = Q_1^4 - \tfrac{3}{14}Q_1^2 + \tfrac{3}{560},$$

and so on. Now these are, in fact, Legendre polynomials defined on the interval $-\frac{1}{2}$ to $\frac{1}{2}$ instead of the more usual -1 to $+1$. The polynomials and this limiting property are, I believe, due to Tchebycheff (1864).

6. For my present purpose it will be more convenient to use Legendre polynomials themselves. Consider a time series $u(t)$ defined (by change of scale of necessary) on the interval -1 to $+1$. We suppose the points of observation so close together as to allow summation to be replaced by integrals in this interval. We fit a polynomial

$$u = a_0 + a_1 P_1 + \ldots + a_r P_r, \tag{5}$$

where P_j is the Legendre polynomial of order j. In view of the orthogonality property

$$\int_{-1}^{1} P_j(x) P_k(x)\, dx = 0 \qquad (j \neq k)$$
$$= \frac{2}{2k+1} \qquad (j = k), \tag{6}$$

we have $a_j = \frac{1}{2}(2j+1)\sum_t u_t P_j$. Moreover

$$P_j(0) = \frac{(-1)^{\frac{1}{2}j}}{2^j}\binom{j}{\frac{1}{2}j}$$

and hence

$$u(0) = \sum_{j=0}^{r} \frac{(-1)^{\frac{1}{2}j}}{2^j}\binom{j}{\frac{1}{2}j}\tfrac{1}{2}(2j+1)\sum_t u_t P_j. \tag{7}$$

Only terms of even order survive in this summation. Taking $r = 2p$ we then have

$$u(0) = \sum_{j=0}^{p} \frac{(-1)^j}{4^j}\binom{2j}{j}\tfrac{1}{2}(4j+1)\,\Sigma P_{2j} u. \tag{8}$$

Thus the weight function attached to the variable u to give the trend value $u(0)$ is

$$\sum_{j=0}^{p} \frac{(-1)^j}{4^j}\binom{2j}{j}\tfrac{1}{2}(4j+1) P_{2j}. \tag{9}$$

I proceed to show that this tends to the $(2p+2)$th derivative of a normal frequency function. Why this should be so is still something of a mystery to me. The sum in (9) is, it should be noticed, an ordinary arithmetic sum, not a convolution of random variables for which a central limit effect might be expected to generate some kind of normality.

7. The moment of order $2k$ of (9) is given by

$$\int_{-1}^{1} \sum_{j=0}^{p} \frac{(-1)^j}{4^j}\binom{2j}{j}\tfrac{1}{2}(4j+1)\, t^{2k} P_{2j}(t)\, dt. \tag{10}$$

Odd-number moments, of course, are zero. Now it is known (cf. for example, Edwards, 1922, vol. 2, p. 902) that

$$\int_{-1}^{1} t^{2k} P_{2j}\, dt = 0 \quad (j > k)$$
$$= 2^{2k+1}\{(2k)!\}^2/(4k+1)! \quad (j = k)$$
$$= \frac{k(k-1)(k-2)\ldots(k-j+1)\ldots(j \text{ factors})}{(k+\frac{1}{2})(k+\frac{3}{2})\ldots(k+j+\frac{1}{2})\ldots(j+1 \text{ factors})} \quad (j < k). \tag{11}$$

We then have, for the $2k$th moment of the weighting function,

$$\mu_{2k} = \sum_{j=0}^{p} \frac{(-1)^j}{4^j}\binom{2j}{j}\tfrac{1}{2}(4j+1)\frac{k(k-1)\ldots k-j+1}{(k+\frac{1}{2})\ldots(k+j+\frac{1}{2})} \quad (j \leqslant k). \tag{12}$$

μ_0 is unity. For $p = 0$ we find

$$\mu_{2k} = \frac{1}{2k+1}.$$

For $p = 1$

$$\mu_{2k} = -\frac{3(k-1)}{(2k+1)(2k+3)}.$$

For $p = 2$

$$\mu_{2k} = \frac{15(k-1)(k-2)}{2(2k+1)(2k+3)(2k+5)}.$$

It is then easy to show by induction that

$$\mu_{2k} = (-1)^k \frac{3.5\ldots(2p+1)(k-1)(k-2)\ldots(k-p)}{(2k+1)(2k+3)\ldots(2k+2p+1)} \quad (p>0).$$

Thus the moment generating function of the weight function (9) is

$$1+\sum_{k=1}^{\infty}(-1)^k \frac{3.5\ldots 2p+1(k-1)\ldots(k-p)}{(2k+1)\ldots(2k+2p+1)}\theta^{2k}. \tag{13}$$

Apart from the initial value of unity the first terms vanish for $k = 1, \ldots, p$.

8. Consider the successive terms in (13). If we subtract the first term, unity, corresponding to u_t itself, we obtain the moment generating function of the residual as

$$(-1)^{p+1}\frac{3.5\ldots(2p+1)\ldots p!\,\theta^{2p+2}}{2p+1\ldots 4p+3}\left[1-\frac{(p+1)(2p+3)}{4p+5}\theta^2+\frac{(p+1)(p+2)(2p+3)(2p+5)}{2!(4p+5)(4p+7)}\theta^4-\text{etc.}\right].$$

The term in square brackets is, for large p, approximated by

$$1-\frac{p\theta^2}{2}+\frac{1}{2!}\left(\frac{p\theta^2}{2}\right)-\text{etc.} = e^{-\frac{1}{2}p\theta^2}.$$

Thus, for large p the moment generator of the weight function is approximately

$$\text{constant.}\,(p\theta^2)^{p+1}e^{-\frac{1}{2}p\theta^2}. \tag{14}$$

Now this, with a suitable scale, is the m.g.f. of $(d/dx)^{2p+2}e^{-\frac{1}{2}x^2}$ the normal function multiplied by the Tchebycheff–Hermite polynomial of order $2p+2$. And this (continuous) quantity is approximately represented by the discontinuous quantity Δ^{2p+2} acting on an approximately normal function. Hence, if we fit a polynomial of order $2p$ the weighting function of the residual is approximated by the $(2p+2)$th difference of a normal function. This is the theorem.

9. As a numerical illustration I took the formula for fitting a quartic or quintic to 21 points, which has the residual

$$\frac{1}{260{,}015}\Delta^6[11{,}628;\ 63{,}308;\ 192{,}423;\ 426{,}258;\ 759{,}003;\ 1{,}135{,}134;\ 1{,}460{,}449;\ 1{,}591{,}294;\ \ldots]. \tag{15}$$

Taking the figures in square brackets (which total 9,687,700) as the ordinates of a grouped distribution I find for the second moment about the mean, with Sheppard's correction

$$\mu_2 = 5{\cdot}638{,}402 - 0{\cdot}083{,}333 = 5{\cdot}555{,}069.$$

The proportional ordinates of (15) are compared with the ordinates of a normal curve with this variance as follows:

Deviation from mean	Observed	Normal
0	0·387	0·399
1	0·355	0·365
2	0·276	0·278
3	0·185	0·177
4	0·1037	0·0945
5	0·0468	0·0420
6	0·0154	0·0156
7	0·00283	0·00485

For p as low as 2 this seems a very fair fit.

10. The approximation (14) to the m.g.f. of the weighting function of the residual and the corresponding expression for the function itself may also be used to provide information about the autocorrelations of residual terms. Suppose that the original series u consisted of a polynomial p plus a random term. On the scale for which our weighting function is proportional to $D^{2p+2}e^{-\frac{1}{2}x^2}$ the autocovariance of the residuals with lag c is then proportional to

$$\int_{-\infty}^{\infty}(D^{2p+2}e^{-\frac{1}{2}x^2})(D^{2p+2}e^{-\frac{1}{2}(x-c)^2})\,dx$$

which, by partial integration, reduces to

$$\int_{-\infty}^{\infty} \{D^{4p+4} e^{-\frac{1}{2}x^2}\} \{e^{-\frac{1}{2}(x-c)^2}\} dx$$

$$= \int_{-\infty}^{\infty} H_{4p+4}(x)\, e^{-\frac{1}{2}\{2(x-\frac{1}{2}c)^2 - \frac{1}{2}c^2\}}$$

$$= e^{-\frac{1}{4}c^2} \int_{-\infty}^{\infty} H_{4p+4}(x)\, e^{-(x-\frac{1}{2}c)^2} dx. \tag{16}$$

The generating function of $H(x)$ is given by

$$\sum_{j=0}^{\infty} \frac{t^j H_j(x)}{j!} = \exp(tx - \tfrac{1}{2}t^2).$$

Multiplying by $\exp\{-\frac{1}{2}(x-\frac{1}{2}c)^2\}$ and integrating over x we find, on some re-arrangement, for the expression (16) a constant multiplied by

$$\exp(-\tfrac{1}{4}c^2)\, H_{4p+4}\left(\frac{c}{\sqrt{2}}\right).$$

For $c = 0$ this is equal to

$$\frac{(4p+4)!}{(2p+2)!\, 2^{p+2}}$$

and hence the autocorrelation function is

$$\rho(c) = \frac{(2p+2)!\, 2^{2p+2}}{(4p+4)!} \exp(-\tfrac{1}{4}c^2)\, H_{4p+4}\left(\frac{c}{\sqrt{2}}\right). \tag{17}$$

The correlogram is accordingly the function (17) graphed against c, and has the appearance of a damped oscillator. The power spectrum, being the Fourier inversion, will be of type $(c^{4p+4} e^{-c^2})$, a bell-shaped curve falling rapidly towards zero.

Similar methods can clearly be used when the original series consisted of a polynomial plus an element which was itself autocorrelated.

REFERENCES

TCHEBYCHEFF, P. L. (1864). Sur l'interpolation. *Zapiski Akademii Nauk.* **4**, Supplement number 5. Collected Works, **1**, 539.

EDWARDS, J. (1922). *The Integral Calculus.* London: Macmillan and Co.

Unbiased estimation of a set of probabilities

BY D. E. BARTON

University College London

1. The recent paper by Blythe & Curme (1960) has drawn attention to the possibility of unbiased estimation of polynomial functions of a parameter of a distribution with particular reference to the notion of completeness introduced by Lehmann & Scheffé (1950). (And it is worthy of notice, in view of what follows, that Stuart's (1955) discussion of his anomaly involves the unbiased estimation of a quadratic in the Binomial p parameter.) This has suggested the problem of obtaining unbiased estimators (u.b.e.'s) of population probabilities which are functions of the maximum-likelihood estimators of the parameters concerned. It is envisaged that these 'population values' will be used not only for the graduation of data but also to provide 'expected group frequencies' from which the χ^2 goodness-of-fit test statistic is to be computed. In § 2 below we treat the particular case of the binomial distribution by direct construction. In § 3 we observe that what we have done may be looked at as an application of Rao's (1945) and Blackwell's (1947) device† and the method is then applied to obtain corresponding results for the Poisson and normal distributions.

2. If we have a set of n independent samples $r_1, \ldots, r_n$ from the binomial distribution

$$P \equiv p(r) = {}^{N}C_r p^r q^{N-r}, \quad r = 0, \ldots, N \quad (q = 1-p),$$

then

$$R = \sum_{i=1}^{n} r_i,$$

has distribution

$$p(R) = {}^{Nn}C_R\, p^R q^{N-n-R}.$$

† This device was first *used* by Rao (1945); it is implicit in equation (3·7) and elsewhere in that paper. Blackwell emphasizes that it provided a *method* of *construction* of an estimator function.

It follows that, for integral values of r

$$\mathscr{E}\{R^{(r)}(Nn-R)^{(N-r)}\} = (Nn)^{(N)}p^r q^{N-r},$$

where $$R^{(r)} = R(R-1)\dots(R-r+1), \quad \text{etc.}$$

Thus if $$\hat{P} = {}^{R}C_r\,{}^{Nn-R}C_{N-r}/{}^{Nn}C_N,$$

Then for all r $$\mathscr{E}(\hat{P}) = P.$$

It is worth noting that one proof of the result that $\hat{P}$ is the *unique* function of R which is an u.b.e. of P is by appeal to Lehmann & Scheffé's result that the distribution of R is complete.

3. More generally let us consider a random variable x whose distribution function depends on a set of parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_\pi)$. If a set of n values $x_1, \dots, x_n$ are randomly and independently drawn from this and their likelihood admits the sets of statistics $\mathbf{t} = (t_1, \dots, t_{\pi'})$, with $\pi \leqslant \pi' < n$, which are collectively sufficient for $\boldsymbol{\theta}$, then let us take an interval $[X_1, X_2]$ and search for an u.b.e., purely a function of $\mathbf{t}$, of the function

$$P \equiv P\{X_1 \leqslant x \leqslant X_2\}$$

which is a function purely of the unknown $\boldsymbol{\theta}$ and the two specified numbers X_1, X_2. The application of the Blackwell–Rao device is to observe that if m is the number of $x_1, \dots, x_n$ which fall in $[X_1, X_2]$ then

$$\hat{P} = \frac{1}{n}\mathscr{E}(m|\mathbf{t})$$

is an u.b.e. of P (since m/n is so, overall). Further, since

$$\mathscr{E}(m|\mathbf{t}) = nP\{X_1 \leqslant x \leqslant X_2|\mathbf{t}\},$$

where x is any of $x_1, \dots, x_n$, then $$\hat{P} = P\{X_1 \leqslant x \leqslant X_2|\mathbf{t}\}.$$

Some examples will illustrate the method.

(*a*) If x is a binomial variable of unknown parameter p and the interval contains just one integer value, r say, then we obtain the expression for $\hat{P}$ already given in § 2.

(*b*) If x is a Poisson variable of unknown mean λ, the sample mean, $\bar{x}$ say, is sufficient for λ and if the interval contains just one integer value r, then

$$\hat{P} = \frac{(n\bar{x})^{(r)}}{r!\,n^r}\left(1-\frac{1}{n}\right)^{n\bar{x}}$$

(which is easily seen *a priori* to have the expectation $e^{-\lambda}\lambda^r/r!$ as it should). A similar result follows in the same way for a truncated Poisson distribution.

(*c*) If x is normal with unknown mean μ and known standard deviation (taken for convenience to be unity), then $\bar{x}$ is sufficient for μ and, since the joint distribution of $(x, \bar{x})$ has the bivariate normal form $N(\mu, \mu, 1, 1/\sqrt{n}, 1/\sqrt{n})$, then conditionally on $\bar{x}$, x has the normal distribution $N\left(\bar{x}, \sqrt{\frac{n-1}{n}}\right)$. Therefore

$$\hat{P} = \int_{X_1}^{X_2} \frac{\exp\left\{-\frac{n(x-\bar{x})^2}{2(n-1)}\right\}}{\sqrt{(2\pi(n-1)/n)}}\,dx$$

is an u.b.e. of

$$P = \int_{X_1}^{X_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}\,dx.$$

(*d*) In the same way, if x has the normal distribution $N(\mu, \sigma)$ and we use the sufficient estimators $\bar{x}$ and

$$s = \sqrt{\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}}$$

of μ and σ, then the integral over $[X_1, X_2]$ of the distribution of x conditional on $\bar{x}$ and s, viz.

$$P = \int_{X_1}^{X_2} \frac{\sqrt{n}}{(n-1)\,B(\frac{1}{2}, \frac{1}{2}n-1)}\left\{1-\frac{n}{(n-1)^2}\left(\frac{x-\bar{x}}{s}\right)^2\right\}^{\frac{1}{2}n-2}\frac{dx}{s}$$

is an u.b.e. of

$$P = \int_{X_1}^{X_2} \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx.$$

[The integrand in $\hat{P}$ is to be interpreted as zero for $|x-\bar{x}| > s(n-1)/\sqrt{n}$.]

4. When the method of the present note is used to estimate the expected group frequencies for use with the χ^2 test it will be seen that it is a relatively easy matter to get the conditional moments of χ^2 explicitly, and so to investigate the small sample application of the loss of degrees of freedom rule, but

this we will leave for a future note. When the data are 'grouped up' the estimators remain u.b.e.'s but it will be noted that $\bar{x}$ and s are *not* the maximum likelihood estimators in the case of grouped normal frequencies.

References

Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.* **18**, 105–10.

Blythe, C. R. & Curme, G. L. (1960). Estimation of a parameter in the classical occupancy problem. *Biometrika*, **47**, 180–5.

Lehmann, E. L. & Scheffé, H. (1950). Completeness, similar regions and unbiased estimation. *Sankhyā*, **10**, 305–40.

Rao, C. R. (1945). Information and accuracy in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–81.

Stuart, A. (1955). A paradox in statistical estimation. *Biometrika*, **42**, 527–9.

Corrigenda

Biometrika, (1960), **47**, pp. 345–53.

'Small sample behaviour of certain tests of the hypothesis of equal means under variance heterogeneity.'

By R. S. McCullough, J. Gurland and L. Rosenberg

p. 352, Table 3, last line in third column

read ·27 *for* ·74

Biometrika, (1960), **47**, pp. 469–73.

'Some distributions arising in the study of generalized mean differences.'

By T. A. Ramasubban

p. 470, equation (1·4) *read* $\sum_{s=1}^{r+1}$ *for* $\sum_{s+1}^{r+1}$;

equation (3·3) *read* $\sum_{d=0}^{n}$ *for* $\sum_{d=0}^{r}$;

equation (3·5) *read* $\sum_{j=0}^{[\frac{1}{2}(r-2)]}$ *for* $\sum_{j=0}^{[\frac{1}{2}(r-1)]}$ (twice).

p. 471, equation (3·7), 1st line, left-hand side *read* ${}_n\Delta_{2r}$ *for* ${}_n\Delta_{2n}$,

2nd line, right-hand side *read* ${}_{(n-1)}\Delta_0$ *for* $(n-1)\,\Delta_0$.

Biometrika, (1960), **47**, pp. 433–37.

'Tables for making inferences about the variance of a normal distribution.'

By D. V. Lindley, D. A. East and P. A. Hamilton

p. 434, five lines from the end of §1

for σ^2 *read* σ^{-2}.

Biometrika (1961), **48**, 1 *and* 2, *p.* 231
Printed in Great Britain

Reviews

The Theory of Storage. By P. A. P. MORAN. London: Methuen and Co. Ltd.; New York: John Wiley and Sons, Inc. 1959. Pp. 111. 13*s.* 6*d.*

This is the first of the series of Methuen Monographs on Applied Probability and Statistics, edited by Prof. Bartlett; it has the same attractive format (and reasonable price) as, for example, the biological series, and the mathematical printing is of good standard.

After introducing some necessary (mainly mathematical) definitions, notations, ideas and techniques the book consists largely of a development of the distribution theory arising from the various models (of stochastic process type) which have been proposed for dealing with inventory and dam theory. It closes with two short chapters: one on Monte Carlo methods (of the direct simulation type) and the other on the choice of best or good Rules of Release (of stock or water, as the case may be) by means of such examples as that where several dams in series, with electric generators of different efficiencies, have to supply a given amount of power with minimal expected loss due to overflow. The Rule of Release problem is shown in the examples to reduce to one in Linear Programming or Calculus of Variations, and methods of solution of these are discussed or referred to.

No problems of statistical inference, either of estimation or testing, in relation to actual realizations of the stochastic processes, are considered and, indeed, no actual data is given although the models (and their shortcomings) are discussed qualitatively in relation to real situations and, also, questions of testing the assumptions of the models from ancillary data (e.g. the independence of successive annual river-flows) and estimation of their parameters *are* treated (by the usual methods of elementary statistics). This is not a criticism of the author who has served the reader well in digesting some two score papers on storage problems but merely reflects the fact that the theory is still in the primitive 'model-building' stage. He points out that, in any event, such necessary data as the annual flow of Australian rivers is apt to be too sparse for any exacting conclusions to be drawn and he is careful to warn against stretching a model too far. Further, the models have been commonly set up, not to describe any existing situation, but to answer the question of what would happen if various rules of restocking and release were adopted.

As is the rule in such theory (which is closely similar to that of queuing) it is relatively easy to set up the Kolmogorov or equivalent equations and only slightly more difficult to pick out any embedded Markov processes. The main task is the exact or approximate solution of these equations. Again, as usual, these are, for the most part, for probabilities of the limiting 'equilibrium' probabilities though the author gives reasons, not always entirely convincing, to suggest that these will be adequate for the real situations envisaged.

In short, this book is an admirably clear introduction to the probability theory of inventory and dam processes, assuming only a knowledge of elementary statistical and stochastic process theory and a corresponding level of mathematics. Application is largely to the comparison of inventory or dam schemes (by deduction of aspects of their behaviour) rather than to calibration of existing processes.

D. E. BARTON

Mathematical Methods in the Theory of Queueing. By A. Y. KHINCHINE. London: Charles Griffin and Co. Ltd. 1960. Pp. 120. 32*s.*

Khintchine's book has been translated from the Russian by Messrs D. M. Andrews and M. H. Quenouille and is one of Griffin's series of statistical monographs. In general terms the book is an elegant, original and rigorous development dealing largely with certain results of Erlang and Palm. Only a few rather special problems are treated; the recent flood of mathematical work on queueing is not dealt with. In fact, the book is a specialist work for mathematicians who are not experts in stochastic processes and is not a text-book for operational research workers; the latter fact is of course precisely conveyed in the title.

The book is in three parts. Part I, which occupies about half the book, deals solely with the properties of 'streams', i.e. of the series of arrivals which form the input into the queue. Chapter 1 derives some simple properties of the Poisson process or, as it is called here, the simple stream. The cases of

constant and variable intensity are dealt with. Chapter 2 is devoted mainly to proving that the only stream that satisfies certain conditions including the 'absence of after-effects' is what may be called an aggregated Poisson stream, i.e. consists of a Poisson process of 'points' together with a distribution determining the number of arrivals per point. Chapter 3 continues these ideas. Chapter 4 deals with streams with limited after effects. This is what in the British and American literature is called the 'general independent' arrival pattern, i.e. the intervals between successive arrivals are independent identically distributed random variables. Chapter 5 is devoted to a careful proof of the result that the superposition of a large number of independent processes leads, under very general conditions on the individual processes, to a Poisson process. Part I is thus an essay on 'point processes', and not concerned specifically with queueing.

The second half of the book deals with queueing processes as such, part II with systems with losses and part III with systems allowing delay. In part II the first chapter, Chapter 6, deals with a random stream of calls entering a system with n lines, the distribution of service-time being exponential. A call arriving when all servers are busy is lost. The formulation of the equilibrium equations is very carefully discussed and the ergodic interpretation of the equilibrium probabilities proved. Incidentally the open problem mentioned at the top of p. 69 has been solved by Karlin (in *Studies in the Mathematical Theory of Inventory and Production*, editors J. Arrow *et al.*, Stanford University Press). Chapter 7 is concerned with the system with infinitely many lines and Chapter 8 with some interesting results concerning the series of events entering the rth line, when calls are assumed to occupy the line with lowest serial number amongst those free at the moment of entry. In the last three chapters, part III, systems with waiting are discussed, the three chapters dealing respectively with exponential service-time, and with the constant service-time single-server queue, and with the general service-time, single-server queue.

The mathematical style of the book is of the highest quality. Results are precisely stated, but the mathematics is kept to an elementary level and irrelevant abstractness avoided. The author was, of course, until his recent death, one of the most eminent of the Soviet school of probabilists and his interest in queueing problems dates from the early 1930's.

The translation is on the whole very clear indeed. A few minor points: the T. Fray mentioned in the bibliography and once or twice in the text may well be T. Fry, the American author of the well-known book on probability; at the top of p. 19 there seems to be an unnecessary 'not', and there is an obvious slip in the final paragraph of p. 60.

D. R. COX

Regression Analysis. By E. J. WILLIAMS. London: Chapman and Hall; New York: John Wiley and Sons, Inc. 1959. Pp. 214. 60*s*.

'This book is addressed primarily to research workers in the experimental sciences. The problems with which it deals have arisen from consultation with such workers,...'—so states the author in his Preface (dated September 1959). In at least one quarter this book has been dismissed rather summarily in terms of '...the point of view is that which existed before the advent of statistical decision theory'. It is true that the comments given by the author (§ 2.9) on the problem of allocating the values of the independent variable are somewhat brief and one might have expected some reference to the few papers which did exist on this topic—by Elfving, Chernoff, Hoel and Guest. But it would have been difficult for the author to have included the recent and substantial paper by Kiefer & Wolfowitz [*Ann. Math Statist.* (1959), **30**, 271–94]. Hence this is really not a satisfactory ground on which to pass judgement. The point that gives this reviewer cause for some concern is the rather sweeping title in relation to the material actually found in the book; although this material does apparently fulfil the standpoint of the author as given in the Preface.

The eleven chapters of this book are intended to pivot round the seventh on Analysis of Covariance. The Introduction, dealing with some general considerations relating to tests of significance, is followed by three chapters on Linear Regression, Multiple and Polynomial Regression and Regression Equations requiring Iterative Calculation. The fifth and sixth chapters discuss Choice among Regression Formulas and Estimation from the Regression Equation. The four chapters following the one on covariance—which itself is introduced as a variant of multiple regression—are intended to help with the treatment of important multivariate problems. In the eighth chapter is a Treatment of Heterogeneous Data and this is followed by three chapters on Simultaneous Regression Equations, Discriminant Functions and Functional Relations. A list of seventy-five references covering a period from the early 1930's to 1958 together with an Index in the standard form complete the book.

There are a number of interesting features about Dr Williams's book. He uses a fiducial argument throughout which induces a rather 'cathedral-like' atmosphere into the presentation of some sections. Moreover, because the reader is assumed (*vide* Preface) to be familiar with—or have ready access to—general statistical methods, some of the drafting is *ex cathedra* to the point of being brutal (to the experimental scientist in mind). Topics covered in Chapter 4 include exponential and hyperbolic regression and the author may regret that a group of important papers on the former topic should have appeared just as his book was in Press. Chapter 5 ('Choice among Regression Formulas') is not quite that which it may seem. In fact it deals with how different independent variables and forms of regression functions may be compared and not the choice among a set of alternatives of a regression formula appropriate for the description of a collection of data. In the latter part of the book the reader will find discussions on tests of concurrence and of proportionality of regression lines; comparison of regressions derived from series which are cross-correlated; simultaneous regression equations in the field of chemistry and the method of instrumental variables applied to a problem of metal surface finishing.

All this is most valuable and within the declared aims of the author. But does it justify the broad title under which the book is being sold? There are many well-known regression problems of importance to research workers which are not included in this book but which would reasonably be expected to be found in a monograph taking the title in question. As the book literature of statistical theory and method changes its emphasis from the general treatise to the monograph, an author will need to describe his work more closely and the publisher accept the (probably extended) title although it may appear superficially less attractive on the advertising material. Nevertheless, Dr Williams is to be thanked for making available his experiences in this field of statistics and congratulated on producing a co-ordinated review of regression methods which can be recommended to those for whom he writes and others with similar problems.

WM. R. BUCKLAND

Elementary Decision Theory. By HERMAN CHERNOFF and LINCOLN E. MOSES. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1959. Pp. xv + 364. 60*s*.

The authors suggest as a first course in statistics from the decision making point of view an introductory chapter, and then chapters on data processing, probability and random variables, utility and descriptive statistics, uncertainty due to ignorance of the state of nature, the computation of Bayes strategies, and introduction to classical statistics. The book ends with a valuable chapter on models, and two rather condensed but comparatively standard chapters on hypothesis testing and estimation and confidence intervals. There are eight tables, eighteen appendices, a partial list of answers to exercises, and an index.

It is easy to criticize the book. The reader is expected to need a description of graph paper, and yet in the introductory chapter, seven pages suffice for the introduction (in the discussion of an example) of the concepts of available actions, states of nature, losses, an experiment, observations, frequency of responses, strategies, average loss, action probabilities, domination of strategies, admissibility, and the minimax rule. All this, before any mention of probability or expectation. Throughout the book, concepts are introduced and discussed in terms of the curious family and communal lives of the inhabitants of the Phiggins towns, the mythical nature of which becomes the more evident as the book progresses.

There seems to have been a difference of opinion between the authors, who claim to be writing a book about statistics, and the publishers, who call it a book of related interest to statisticians. Certainly it would be impossible to obtain an adequate grasp of what is usually called elementary statistics from this book alone. The normal mean test and the t test are both dismissed in a single exercise; the χ^2 distribution occurs only in the solution of an exercise; the existence of non-parametric methods is barely mentioned. Nevertheless, for students with some more orthodox grounding, this book, by obvious enthusiasts, may prove stimulating.

C. L. MALLOWS

Individual Choice Behaviour. By R. DUNCAN LUCE. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1959. Pp. xii + 153. 48*s*.

This monograph contains an axiomatic approach to the probability models used in modern experimental psychology. It is thus concerned with how an individual actually does behave in uncertain situations, rather than with how an idealized 'rational' individual would behave. The basic axiom of choice behaviour is simply that in cases of imperfect discrimination, conditional choice probabilities can be defined in the usual way; in this context, this is not a tautology, since two distinct experiments would be required to verify the axiom. The consequences of this axiom are worked out with complete rigour and in great detail; it is shown to imply the existence of a ratio scale over all the possible choices. Further chapters describe applications of the basic theory to psychophysics, to utility theory (where a theory of decomposable preference structures is developed which leads to some remarkable predictions), and to learning theory (where three distinct learning models are constructed). The book ends with an excellent short chapter of summary and conclusions, three mathematical appendices, a list of open problems, bibliography and index.

As an exercise in axiomatics, the book is a model. With one exception, every assumption made is stated with full precision and is carefully discussed; experiments for testing the theory are described and proposed. The exception occurs in the section on response bias, where an assumption similar to that of 'no interaction' in a factorial experiment is made, and is subsequently used without discussion in the construction of models for signal detectability experiments.

The author has 'strong'reservations' about the detailed accuracy of the expected utility hypothesis; for statisticians who are not psychologists, it is in his remarks on this subject, and in the appeal of a careful discussion of a basic probability theory in an unfamiliar context, that the value of the book will mainly lie.

C. L. MALLOWS

Principles and Procedures of Statistics (with special reference to the biological sciences). By R. G. D. STEEL and J. H. TORRIE. London, New York and Toronto: McGraw-Hill Publishing Co. 1960. Pp. xvi + 481. 81*s*. 6*d*.

This is a 'cook-book' and a good one for the use of biologists who wish to acquire the elements of biometry. The statistical techniques are presented with the minimum of algebraic manipulation, the procedures to be adopted are clear and the whole is illustrated with many actual illustrations, most of which have the merit, to the reviewer at least, of being fresh material.

Biologists, no matter from which side of the Atlantic they come, are timorous about algebra and even about algebraic notation, a trepidation which they often cover up with an aggressive disdain for abstract thought. It is difficult to see how Messrs Steel and Torrie could have used less symbolism than they have, but the average biologist will not find the book easy to read. For those who are forced to apply the techniques set out, understanding will come through hard work. For those who merely want to learn a smattering of a new subject something easier will need to be sought after.

The scope of the book is wide. There are the usual descriptive statistics, probability, distributions of sample criteria, experimental design, analysis of variance (one-way and multi-way classifications), linear regression and correlation, more analysis of variance including factorial experiments, splitplots and unequal subclass numbers, multiple and partial regression and correlation, analysis of covariance, nonlinear regression, χ^2 criterion and its various applications, the hypergeometric, binomial and Poisson distributions, non-parametric statistics and stratified sampling. The first chapter on the definition of statistics and its history reads a little oddly and could with advantage have been omitted. The readers of this book are not going to ask themselves what statistics is (or are) but merely how the methods are used.

The book can be recommended.

F. N. DAVID

The Statistical Basis of Quality Control Charts. By S. K. EKAMBARAM. India, London and New York: Asia Publishing House. 1960. Pp. x + 96. 16*s*. 6*d*.

The author of this short book sets himself a stiff task. His opening sentence in the preface reads 'This book is written for the busy executive in the higher levels of Management who desires to have the general picture of the technicalities of quality control in industry, but has no time to go into the

details of quality control work.' Some idea of how the author tries to achieve this can be gained from the selection of topics discussed, which fall into three groups. The first third of the book deals with basic statistical theory such as frequency distributions, variation, moments, measures of skewness and so forth. The next third deals with control charts based on the mean and range of small samples, whilst the final third deals with charts based on the observed fraction defective in samples.

For the executive, the book's great weakness is the unnecessary amount of detailed algebra that is included. The derivation of the moments about the mean from crude moments is dealt with extensively, and the first four theoretical moments of the normal, binomial and Poisson distributions are all derived and studied in considerable detail. This approach tends to dull the appetite for quality control. Although a lot of this algebra could be skipped in reading the book, the continuity would be broken. If, however, the executive does persevere there are some good case studies, five in number, which bring out clearly the principles involved and it is a pity that the book was not primarily written around these studies, rather than around the three basic distributions mentioned above. A brief discussion of how to select a quality control scheme, how to organize a quality control section, and how to assess its efficiency, would also have been helpful to the executive.

The practising statistician will only find this book of very general interest. There is no mention of double sampling or sequential schemes, modified limits, demerit schemes or the more modern forms of continuous process schemes based on cumulative charts. All charts discussed use the American system of 3-sigma limits, and warning limits are not used. No tables giving the constants required for control limits are included in the book. Finally, the operating characteristic is only briefly touched upon in the last chapter, and there is no detailed discussion of how to set about choosing an appropriate scheme, or how to assess the economics of alternative schemes.

P. G. MOORE

An Introduction to Mathematical Statistics. By H. D. Brunk. Massachusetts: Ginn and Company. 1960. Pp. xi + 403. $7.00.

The author states that his book was written for use in a one-semester, 3 hr. course which is meant as an introduction to mathematical statistics. Prerequisites include a knowledge of the calculus—and the author might have added algebra—but nothing in probability and statistics, which means that the student can start *de novo*. This states the level adequately for American readers. As far as the English student is concerned he will need Pure Mathematics at G.C.E. A-Level at a fairly good standard. The author has added starred sections and chapters to his original planned course. These will require more mathematical knowledge than is necessary for the unstarred sections. The book is in two parts, probability and statistics.

In probability we have elementary probability spaces, general probability spaces, random variables, combined random variables and expectation techniques. Under statistics we find random sampling, law of large numbers (including Tcheby-chef), estimation, central limit theorem, confidence intervals and testing hypothesis, decision theory, regression, distributions of various statistical criteria, experimental design, quality control and distribution-free methods. The whole is topped up with a number of useful statistical tables. The level of exposition is adequate and there are some numerical illustrations.

F. N. DAVID

An Introduction to Mathematical Statistics. By R. V. Hogg and A. T. Craig. New York and London: The Macmillan Company, New York. 1960. Pp. ix + 245. 47*s*.

The authors state that this book is intended for a two-semester course for undergraduate students with some mathematical preparation. They indicate which sections will be found useful if it is desired to cut the course down to one semester. From the English point of view the students will require little more than advanced level mathematics and the book will certainly be within the compass of the first year general student.

Probability and statistics are intermingled throughout in a way which comes naturally. The topics treated are random variables and expectations, the binomial, Poisson, normal and gamma distributions, density functions, transformations of variables, estimation sufficiency, and stochastic independence, limiting distributions, distribution-free problems, tests of statistical hypotheses, and multivariate distributions. Tables of the normal, χ^2, t and F distributions are given.

The mathematical exposition is clear and the seeker for light on the mathematics of statistics will find this book meets his needs admirably.

F. N. DAVID

Genetical Research. Vol. 1, no. 1. London and New York: Cambridge University Press. 1960. Pp. 172. Subscription (three parts annually) £5 or $17.50.

The rapid growth of genetics in recent years has necessitated changes in existing journals. *Genetics* has lately considerably increased in size, and *Heredity* this year appears in two volumes rather than one preparatory to an increase in girth. The need of additional vehicles for publication has now led to the emergence of this new journal. The editorial board includes C. H. Waddington (chairman), Charlotte Auerbach, H. Grüneberg, W. Hayes, D. Lewis, G. Pontecorvo and A. Robertson, with E. C. R. Reeve as Executive Editor. The journal is open to authors of all nationalities, but papers must be written in English. The twelve papers which comprise the first number range over almost the whole field of genetics and deal with organisms as diverse as fungi (*Neurospora*, *Coprinus*), protozoa (*Paramecium*), *Drosophila*, the mouse and the rabbit. As was to be expected of the Cambridge University Press, the production of the new journal is of high quality.

H. GRÜNEBERG

La Biométrie. Que sais-je. (Le point des connaissances actuelles, no. 871.) By EUGÈNE SCHREIDER. Paris: Presses Universitaires de France. 1960. Pp. 126.

This small paper-back is a recent addition to the 'Que sais-je?' series. The author's objects are to refute the contention that biometry is a dull, uninteresting science ('une science aride') and to show the non-specialist reader something of the nature and scope of biometrical research.

The contents are divided into four parts, the first of which deals with some elementary statistical procedures, while the remaining three are concerned with their application to biological data. Most of the examples are taken from the human species. There are seven figures and various tables, but the book lacks an index. Somewhat surprisingly the author has made some use (Tables III and IV) of the now rarely used coefficient of variation. In other respects the book is up to date and recent researches are cited, including some by the author.

The section on variability in populations due to heredity and environment, and that dealing with changes in man and illustrated by the increase in height in various countries over a period of years, are of general interest.

S. B. HOLT

Introduction to Health Statistics. By SATYA SWAROOP. Edinburgh and London: E. and S. Livingstone, Ltd. 1960. Pp. 343. 40*s*.

Dr Swaroop, who is a statistician working with the World Health Organization, has used his very full knowledge of the organization, collection and use of health statistics to produce an excellent reference book on the subject. It summarizes well the various recommendations and findings of the Expert Committees, in one volume. Whether, however, it is suitable for use by students is questionable. Its style is not well adapted to easy reading. The text is interspersed with lists and indications for procedures, which are useful on reference and when learning off by heart, but not otherwise. It is, also, in places, difficult to evaluate the dangers, drawbacks and fallacies of certain procedures, particularly in the sections on census procedures and morbidity surveys. The book contains a full account of the types of data the World Health Organization recommends should be collected, with details of the necessary forms, discussion of the use and need for health statistics, methods of collection of the data and how it should be analysed and some details of the necessary calculations and diagrams. The emphasis in this is particularly on the methods recommended and used by W.H.O. and, therefore, of more use to an individual starting a health service *de novo*, rather than working in an existing organization. The appendices contain various international comparisons of birth and death rates, a very full reference list of the methods of vital statistics, and a list of some recent W.H.O. expert committees and international regulations.

W. W. HOLLAND

Special Functions. By EARL D. RAINVILLE. New York and London: The Macmillan Company, New York. 1960. Pp. xii + 365. 82*s*.

The statistics student, who is not a mathematics specialist, is often brought up short in his study of mathematical statistics by various manipulations of the special functions which lie in wait in many branches of the theory. And elucidation from mathematical text-books is not easy because the delineation of the special bit of theory required is usually tied up with a great deal which is irrelevant to the particular purpose in hand. This book, pitched at just about the right level of exposition, is excellent, and can serve as supplementary reading for any searcher after mathematical statistical truths.

The contents cover (briefly), infinite products, gamma and beta functions, asymptotic series, the hypergeometric functions, Bessel functions, generalized and confluent hypergeometric functions, generating functions, orthogonal, Legendre, Hermite, Laguerre and Jacobi polynomials, elliptic and theta functions. There are exercises and examples given at the end of most chapters.

Mr Rainville says he has written his book to be of use to physicists, engineers and chemists. The book can be unreservedly recommended to statisticians also.

F. N. DAVID

The Theory of Matrices. By F. R. GANTMACHER. New York: Chelsea Publishing Company. 1959. Vol. 1. Pp. 374. Vol. 2. Pp. 276. $6.00 each.

This excellent translation from the Russian, made by Prof. K. A. Hirsch, will be welcomed by the large and growing number of mathematicians, scientists, and statisticians who work with matrices. No knowledge of matrix theory is assumed, but previous knowledge of determinants and of the theory of solution of linear equations is essential. This may limit the usefulness of the book to beginners who are interested in learning linear algebra in a strictly logical order (it is, for instance, taken for granted in the book that the space of n-tuples of complex numbers is n-dimensional); there is a tendency in the book to deduce elementary properties of matrices from rather more advanced theory than is necessary (the Cauchy–Binet formula to prove that the rank of AB cannot exceed that of A or B). But for those mainly interested in applications, these pedagogic objections are of little importance: the book really has advanced aims.

The mathematical treatment is everywhere rigorous, lucid, and amply illustrated by numerical examples. Much attention is given to the serious problems of numerical analysis which abound in the applications, for instance, that of effectively computing the characteristic function of a matrix.

The matrix as an expression of a linear operator, both with and without metric notions, is given good prominence. Thus, an exhaustive analytic account of the theory of elementary divisors is supplemented by an illuminating geometric exposition. Many applications discussed relate to differential equations, in particular, to the stability of solutions. The Routh–Hurwitz method for deciding whether the eigenvalues of a matrix have negative real-parts is discussed fully, together with the contributions of Hermite and Lyapunov to this important practical problem (the theory of quadratic and of Hankel forms is here brought in). A chapter on matrices with non-negative elements (in particular, stochastic matrices) includes the remarkable theorems of Perron and Frobenius on the eigenvalues of such matrices. There is an excellent account of matrix equations and of functions of matrices: in these the reduction to Jordan normal form is made to play an important part. A compendious bibliography lists some four hundred titles, including papers as recent as 1959.

H. KESTELMAN

Theory and Solution of Ordinary Differential Equations. By DONALD GREENSPAN. New York and London: The Macmillan Company, New York. 1960. Pp. viii + 148. 38*s*. 6*d*.

With the growth of the 'dynamics' of statistical method commonly but not necessarily correctly lumped together as 'stochastic processes' the interest of the student of mathematical statistics in the theory of differential equations has perforce been sharpened. This book is written for those who know a reasonable amount of advanced calculus but not all of it will be useful to those who seek to acquire mathematical techniques rather than mathematical learning.

The topics covered are first-order equations, linear differential equations of the second and *n*th order, existence theory, linear systems, special functions, approximate solutions, eigenfunctions and Fourier series. The delineation is clear and the book is eminently readable. Exercises are given at the end of each chapter and the answers supplied. It should serve as a useful introduction to the subject for mathematicians. The 'user' student may find it too difficult.

F. N. DAVID

Pronterari per Calcoli Statistici: Tavole Numeriche e Complementi. By SILVIO VIANELLI. Palermo, Italy: Ed. Abbaco. 1960. Pp. xv + 1543. 16,000 lire.

This monumental piece of compilation consists of reprints of *all* tables which have been calculated for or have been found useful in the application of statistical theory, plus a resumé of the theory for which the tables was calculated where it is appropriate. The tables occupy nearly 1100 pages and the exposition a further 400. A few of the tables appear not to have been previously published. The printing is entirely clear and one can only marvel at the energy of the compiler and the labours of the proof readers. All libraries will need to buy this book if only for record purposes.

This being said the reviewer would like to offer a few suggestions which would improve the book from the point of view of the user. Judging from the relative number of statistical papers published in different languages the majority of users will read the English language. It is, therefore, important that the index to both tables and exposition should be also in English and not solely in Italian as it is at present. It is true that symbols are universal esperanto and that Italian is easy to read but the lack of an adequate index may be a stumbling block to many.

The tables range from the utilitarian—powers of natural numbers, random numbers, etc.—to those of modern techniques only just now passing into general use, such as the Kolmogoroff–Smirnoff criterion. The reviewer was puzzled to find a logical grouping. It is, as things are at present, very difficult to find any particular table without having to search through the whole index and this detracts considerably from the general utility of the book. The expository material is divided according to subject matter with a separate index given for each section which is reasonable to use provided one classifies statistical theory in the same way as the author. A similar set of indices for the tables might be possible, with repetition of entry where this was found necessary.

The physical size of this book will make it difficult to handle and one wonders if consideration was given to producing not one volume but say six or seven. Questions of cost will undoubtedly have entered into this, however, and one can have every sympathy with the author. But all these are minor reflexions. The compiler-author is to be congratulated on his enterprise and for those wealthy enough in the statistical world to be able to afford £9. 8*s*. 3*d*. this is a bargain of bargains.

F. N. DAVID

Other books received

Introduction to Probability and Statistics. By HENRY L. ADLER and EDWARD B. ROESSLER. San Francisco and London: W. H. Freeman and Co. 1960. Pp. xi + 252. 20*s*.

Engineering Statistics. By ALBERT H. BOWKER and GERALD J. LIEBERMAN. New Jersey, U.S.A.: Prentice-Hall, Inc. 1959. Pp. 585. 88*s*.

Information and Design Processes. Ed. ROBERT E. MACHOL. New York, Toronto and London: McGraw-Hill Book Company, Inc. 1960. Pp. 185. 46*s*.

Regression Analysis. By R. L. PLACKETT. Oxford: Clarendon Press and Oxford University Press. 1960. Pp. 173. 35*s*.

Introduction to Linear Programming. By WALTER W. GARVIN. New York, Toronto and London: McGraw-Hill Book Company, Inc. 1960. Pp. 281. 68*s*.

The Use of Economic Statistics. By C. A. BLYTH. The Minerva Series, No. 5. 1960. Pp. 249. 28*s*. (cloth), or 22*s*. (board).

Stochastic Population Models in Ecology and Epidemiology. By M. S. BARTLETT. London: Methuen and Co. Ltd. New York: John Wiley and Sons, Inc. 1960. Pp. 90. 12*s*. 6*d*.

Contributions to Probability and Statistics. Essays in honor of Harold Hotelling. Stanford studies in mathematics and statistics. II. Ed. I. OLKIN. California: Stanford University Press. London: Oxford University Press. 1960. Pp. 517. 52*s*.

Statistical Processes and Reliability Engineering. By D. N. CHOROFAS. Princeton, New York, Toronto, London: D. van Nostrand Co. Inc. 1960. Pp. 438. 96*s*.

An Introduction to Statistical Communication Theory. By DAVID MIDDLETON. New York, Toronto, London: McGraw-Hill Book Company, Inc. 1960. Pp. 1140. £9. 14*s*.

Fundamentals of College Algebra. By WILLIAM H. DURFEE. New York and London: The Macmillan Company, New York. 1960. Pp. 250. 31*s*. 6*d*.

Applied Boolean Algebra. An Elementary Introduction. By FRANZ E. HOHN. New York and London: The Macmillan Company, New York. 1960. Pp. 189. 17*s*. 6*d*.

An Introduction to the Mathematics of Medicine and Biology. By J. D. DEFARES and I. N. SNEDDON. Amsterdam: North-Holland Publishing Company. 1960. Pp. 663. 85*s*.

Sampling Methods for Censuses and Surveys. (Third edition.) By F. YATES. London: Charles Griffin and Company. 1960. Pp. 440. 54*s*.

Finite-difference Methods for Partial Differential Equations. By GEORGE E. FORSYTHE and WOLFGANG R. WATSON. London and New York: John Wiley & Sons, Inc. 1960. Pp. 444. 92*s*. or $11.50.

Symposia of the Society for Experimental Biology. No. XIV. Models and Analogues in Biology. Ed. J. W. L. BEAMENT. Cambridge University Press. 1960. Pp. 255. 50*s*.

Stichproben in der Amtlichen Statistik. Ed. GERHARD FÜRST. Stuttgart and Mainz: W. Kohlhammer 1960. Pp. 626. DM 28.

Sequential Medical Trials. By P. ARMITAGE. Oxford: Blackwell Scientific Publications. 1960. Pp. 105. 20*s*.

BIOMETRIKA

BACK ISSUES

Volumes 43–47 (1956–1960)

These may be obtained from the BIOMETRIKA OFFICE, at the following prices, which include packing and postage:

£3. 10*s*. (or $10.00) per wrappered volume.
Bound volumes: £1. 8*s*. (or $4.00) extra per volume.
Binding cases: 14*s*. (or $2.00) each.

Cheques should be made payable to *Biometrika*, crossed 'a/c BIOMETRIKA TRUST' and sent to THE SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, W.C. 1, to whom all orders for subscriptions and author's separates should also be sent (see inside front cover). The Secretary also has available for sale copies of many of the papers and tables published in earlier numbers of *Biometrika* (see pages (iv), (v) and (vi)).

Volumes 20A, 20B, 21–42 (1928–1955)

Messrs Wm. Dawson & Sons, Ltd., have been appointed agents for the sale of these volumes, which may be obtained from the address below at the following prices, *packing and postage extra:*

£5. 5*s*. (or $15.00) per wrappered volume.
Bound volumes: £1. 8*s*. (or $4.00) extra per volume.
Binding cases: 14*s*. (or $2.00) each.
Index (Subjects Vols. 1–37; Authors Vols. 1–43): 6*s*. (or $1.00).

Volumes 1–19 (1902–1927)

Messrs Wm. Dawson & Sons, Ltd., have been granted permission to reprint Vols. 1–19. These are ready for distribution, price £130 bound. Single vols. £7. 10*s*. each. Would librarians and others wishing to have copies please place their orders with:

WM. DAWSON & SONS, LTD.,
16 WEST STREET, FARNHAM, SURREY, ENGLAND

(*All rights reserved*)

BIOMETRIKA, Vol. 48, Parts 1 and 2

CONTENTS

PAGE

M. G. KENDALL. Studies in the history of probability and statistics. XI. Daniel Bernoulli on maximum likelihood 1

F. N. DAVID and C. L. MALLOWS. The variance of Spearman's rho in normal samples 19

E. C. FIELLER and E. S. PEARSON. Tests for rank correlation coefficients. II 29

J. DURBIN. Some methods of constructing exact tests 41

C. R. HEATHCOTE. Preemptive priority queueing 57

J. HAJNAL. A two-sample sequential *t*-test 65

S. NABEYA. Absolute and incomplete moments of the multivariate normal distribution 77

JOHN S. WHITE. Asymptotic expansions for the mean and variance of the serial correlation coefficient . 85

T. H. STARKS and H. A. DAVID. Significance tests for paired-comparison experiments 95

G. S. WATSON. Goodness-of-fit tests on a circle 109

H. T. GONIN. The use of orthogonal polynomials of the positive and negative binomial frequency functions in curve fitting by Aitken's method 115

A. M. W. VERHAGEN. The estimation of regression and error-scale parameters, when the joint distribution of the errors is of any continuous form and known apart from a scale parameter 125

C. L. MALLOWS. Latent vectors of random symmetric matrices 133

H. LEON HARTER. Expected values of normal order statistics 151

FRANK A. HAIGHT. A distribution analogous to the Borel–Tanner 167

W. L. NICHOLSON. Occupancy probability distribution critical points 175

MASASHI OKAMOTO and GORO ISHII. Test of independence in intraclass 2×2 tables 181

MISCELLANEA

COLIN R. BLYTH and DAVID W. HUTCHINSON. Tables of Neyman-shortest unbiased confidence intervals for the Poisson parameter 191

K. C. SREEDHARAN PILLAI and ANGELES R. BUENAVENTURA. Upper percentage points of a substitute *F*-ratio using ranges 195

A. M. WALKER. On Durbin's formula for the limiting generalized variance of a sample of consecutive observations from a moving-average process 197

D. E. BARTON and F. N. DAVID. The central sampling moments of the mean in samples from a finite population (Aty's formulae and Madow's central limit) 199

MAN MOHAN SONDHI. A note on the quadrivariate normal integral 201

J. GANI. On the stochastic matrix in a genetic model of Moran 203

W. J. EWENS. Departures from assumption in sequential analysis 206

J. C. GOWER. A note on some asymptotic properties of the logarithmic series distribution 212

M. ATIQULLAH. On a property of balanced designs 215

M. J. R. HEALY and J. C. GOWER. Aliasing in partially confounded factorial experiments 218

M. G. KENDALL. Studies in the History of Probability and Statistics. XII. The Book of Fate 220

J. C. TANNER. A derivation of the Borel distribution 222

M. G. KENDALL. A theorem in trend analysis 224

D. E. BARTON. Unbiased estimation of a set of probabilities 227

CORRIGENDA 230

REVIEWS

P. A. P. MORAN. The Theory of Storage 231

A. Y. KHINCHINE. Mathematical Methods in the Theory of Queueing 231

E. J. WILLIAMS. Regression Analysis 232

HERMAN CHARNOFF and LINCOLN E. MOSES. Elementary Decision Theory 233

R. DUNCAN LUCE. Individual Choice Behaviour 234

R. G. D. STEEL and J. H. TORRIE. Principles and Procedures of Statistics 234

S. K. EKAMBARAM. The Statistical Basis of Quality Control Charts 234

H. D. BRUNK. An Introduction to Mathematical Statistics 235

R. V. HOGG and A. T. CRAIG. An Introduction to Mathematical Statistics 235

Genetical Research. Vol. 1, no. 1. 236

EUGÈNE SCHREIDER. La Biométrie. Que sais-je 236

SATYA SWAROOP. Introduction to Health Statistics 236

EARL D. RAINVILLE. Special Functions 237

F. R. GANTMACHER. The Theory of Matrices 237

DONALD GREENSPAN. Theory and Solution of Ordinary Differential Equations 237

SILVIO VIANELLI. Pronterari per Calcoli Statistici: Tavole Numeriche e Complementi 238

OTHER BOOKS RECEIVED 239

Printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)

GE
1
19
29
41
57
65
77
85
95
109
115
125
133
151
167
175
181
191
195
197
199
201
203
206
212
215
218
220
222
224
227
230
231
231
232
233
234
234
235
235
236
236
237
237
238
239